Vol. 49 OCTOBER, 1958 No. 5

The Journal of Educational Psychology

CONTENTS

Relationship of Self-Acceptance to Other Variables with Sixth Grade Children Oriented in Self-Understanding 229
PAUL BRUCE

Effects of Training in Alternative Solutions on Subsequent Problem Solving 239
WALTER I. ACKERMAN AND HARRY LEVIN

The Effects of Different Teaching Methods: A Methodological Study. 245
MARVIN NACHMAN AND SEYMOUR OPOCHINSKY

Curricular Differences in Job Incentive Dimensions Among College Students 250
A. W. BENDIG AND EUGENIA L. STILLMAN

Relationship of Intelligence and Social Power to The Interpersonal Behavior of Children 257
ALVIN ZANDER AND ELMER VAN EGMOND

An Approach to the Measurement of Psychological Characteristics of College Environments 269
C. ROBERT PACE AND GEORGE G. STERN

Preschool IQs After Twenty-Five Years 278
KATHERINE P. BRADWAY, CLARE W. THOMPSON, AND RICHARD B. CRAVENS

The Adequacy of "Meaning" as an Explanation for the Superiority of Learning by Independent Discovery 282
BERT Y. KERSH

AMERICAN PSYCHOLOGICAL ASSOCIATION

The Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

LEE J. CRONBACH
University of Illinois
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
HAROLD E. JONES
University of California
IRVING LORGE
Teachers College, Columbia University
DOROTHEA McCARTHY
Fordham University
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
PERCIVAL M. SYMONDS
Teachers College, Columbia University
GEORGE G. THOMPSON
Syracuse University

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
SARAH WOMACK, Editorial Assistant

The Journal of Educational Psychology publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second class postage paid at the Post Office at Baltimore, Md. under the act of March 3, 1879.

Copyright, 1959 by the American Psychological Association, Inc.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 49 October, 1958 Number 5

RELATIONSHIP OF SELF-ACCEPTANCE TO OTHER VARIABLES WITH SIXTH GRADE CHILDREN ORIENTED IN SELF-UNDERSTANDING[1]

PAUL BRUCE[2]

Child Welfare Research Station, State University of Iowa

The purpose of this study was to investigate with sixth grade children some of the relationships between a measure of self-acceptance and other personality variables. The relationships were studied in two groups of *Ss*: one consisting of pupils who had taken part in a learning program designed to develop a more understanding and analytical approach to their own and others' behavior (the experimental group) and one group consisting of pupils who had not undertaken such a program (the control group).

A review of the literature—particularly the writings of Rogers (**15**) and other researchers with a client-centered therapy orientation—suggested that self-acceptance be defined as the congruence between the way a person thinks himself to be (self-concept) and the way he thinks he would most like to be (ideal-self). Accordingly, in this study, a measuring instrument from which Self-Ideal Discrepancy scores were obtained was devised suitable to a sixth grade population.

Within the framework of a self-ideal discrepancy measure of self-acceptance, a child's indication of his ideal-self can be viewed as his expression of how he feels some of his psychological needs (such as the needs for security, personal worth, status, etc.) can be best satisfied. In other words, underlying the conception of ideal-self is the assumption by the child that if he becomes more like his ideal-self, he will be better able to satisfy some of his secondary needs. For example, implicit in a child's wanting "to be better looking," "to have more friends," "to be less sensitive," etc., is the feeling that one's affectional, security, and status needs (among others) would be better satisfied if these ideals were achieved. If a child, then, feels he is quite unlike his ideals or feels he is not making progress towards his ideals, it might be speculated that adequate satisfaction of several of his secondary needs is being blocked in some way, and thus several predictions concerning his behavior might be made.

It would be expected that a child who indicates a marked discrepancy between his self-concept and ideal-self, feeling his need satisfaction blocked, would show evidences of insecurity in his behavior and would yield responses indicating manifest anxiety. Such relationships between measures of self-acceptance and various measures of insecurity and anxiety have been obtained with subjects of college age and beyond (**1, 2, 7, 13, 16**). One of the major purposes of this investigation was

[1] This paper is based on a doctoral dissertation submitted to the Graduate College of the State University of Iowa in 1957. The writer is indebted to Ralph H. Ojemann for his encouragement and guidance throughout the study and to the members of the Preventive Psychiatry Project staff for their invaluable assistance.

[2] Now at San Diego State College.

to see if the relationship between self-ideal discrepancy (as a measure of self-acceptance) and measures of observed insecurity and manifest anxiety held for a group of sixth grade children.

The orientation program which the experimental group of *Ss* had undertaken was designed to help each pupil develop a "causal," analytical orientation towards his social environment. The program was based on the assumption that a child who is provided with the opportunity for understanding some of the many causes underlying his own behavior and the behavior of people about him will be able to make more effective adjustments. Involved in the program was the special training of the teachers and the use of certain special curricular materials, the details of which have been published elsewhere (**9, 11, 12**).

With respect to this investigation, the orientation program provided a group of *Ss* who had been trained in self-understanding, among other things. Several writers, such as Jersild (**6**) and Rogers (**14**), have indicated that an important characteristic of a self-accepting person is self-understanding. The question arises, then, as to whether a child with increased self-understanding is thereby more self-accepting. It might be hypothesized that as a child gains insight into his own motivations and dynamics—that is, gains in self-understanding—he would give evidence of being more self-accepting. Also to the extent that the orientation program helps the individual learn how to work out his daily situations more effectively, he may be helped to make progress towards his ideal thereby reducing the self-ideal discrepancies.

On the other hand, increased self-understanding may not affect self-acceptance, particularly when self-acceptance is defined in terms of self-ideal discrepancy. For example, an individual may see himself some distance from his ideal, but with increased self-understanding, understand some of the reasons for this situation. In such a case, self-ideal discrepancy might remain high even after the orientation program, although the individual would feel better—that is, be less anxious—about it. Thus, the orientation program might serve to reduce anxieties which normally accompany high self-ideal discrepancy without necessarily reducing the discrepancy itself. It appears, therefore, that the orientation program could have diverse effects upon a self-ideal discrepancy measure of self-acceptance.

The specific questions to which answers were sought in this study are: (*a*) Is there a significant difference between *Ss* with high self-ideal discrepancy and those with low self-ideal discrepancy in average scores on a test of manifest anxiety? (*b*) Does this relationship hold for *Ss* who have taken part in an orientation program in self-understanding? (*c*) Is there a significant difference between *Ss* with high self-ideal discrepancy and those with low self-ideal discrepancy in average ratings on an observation scale of insecurity?

Procedure

Measures used in study. A Self-Acceptance Scale was devised and initially administered to several sixth grade classes. The results and items were analyzed qualitatively and quantitatively. The 10 items finally selected for the revised Self-Acceptance Scale, in general, reflected affective characteristics about which individuals in our culture were thought to have substantial feelings. (Only nine of these items were scored since one proved ambiguous in subsequent administrations of the scale.) A more detailed description and a copy of this scale appears elsewhere (**3**).

Examples of typical items are these: "1. This is someone who feels that others don't like him or her—someone whom nobody seems to care about;" and "8.

This person is happy and cheerful most of the time—one who seems to enjoy what he or she does." Some of the items in the scale were adapted from the Social Analysis of the Classroom Inventory developed by the Horace Mann-Lincoln Institute of School Experimentation (**5**). Descriptive statements were utilized in place of the trait-names (used on most other scales of this type) in order to reduce the amount of ambiguity.

The *Ss* were asked to answer four questions about each of the descriptive statements:

1. Would most boys and girls my age like to be like this person? (This question was not scored but was included as a check on the general favorability or unfavorability of each item.)
2. Am I like this person? (Self-Concept)
3. Am I becoming more like this person? (The results obtained by the use of this question were inconclusive and are not discussed in this paper.)
4. Would I like to be like this person? (Ideal-Self)

The *S* answered each question indicating to what extent he felt it was true by checking along a 5-point scale reading: "very much so," "quite a bit," "somewhat," "not very much," and "not at all."

The particular score yielded by this scale which was used in the major analyses reported in this paper is the Self-Ideal Discrepancy score. This score is the discrepancy between the rating in answer to Question 2 (Self-Concept) and the rating in answer to Question 4 (Ideal-Self) summed for the nine items. Test-retest reliability for this score is reported in the next section.

The other measures used in this study included the Children's Manifest Anxiety Scale developed by Castaneda et al. (**4**) and the Kooker Security-Insecurity Rating Scale. The Children's Manifest Anxiety Scale consists of 42 anxiety items which are answered "yes" or "no." The Kooker scale (**8**) gives evidence of children's behavior which is independent of their responses on a paper-pencil type test and which reflects their "typical" behavior patterns. This scale requires a trained observer who is familiar with the child's behavior to rate him on a series of 19 behavior items, as occurring "frequently," "fairly often," or "seldom." In the present study, each classroom was visited for a five-day period and the children rated by one of the two trained observers (who remained ignorant of the nature of this investigation). Both observers practiced rating the same *Ss* with the Kooker scale until they were in substantial agreement with each other.

With the exception of the Kooker scale, the other measures were administered to all classes by the investigator. The teachers were not involved in the measurement aspects of the investigtion and were not aware of the nature of the study.

Subjects of investigation. The subjects in this study were pupils of eight sixth grade classes in different elementary schools located in comparable neighborhoods with respect to socioeconomic status. The four experimental classes had undergone the orientation program described above. Most of the pupils in two of these experimental classes had been in the program for two consecutive years; whereas the pupils in the remaining two experimental classes had undergone the orientation program for the current year only.

At the beginning of the program, four control classes were selected controlling for some of the teacher variability. Thus, for every teacher of an experimental class, a control teacher was selected and matched on the basis of several variables including age, sex, number of years teaching experience, and educational level. Unlike the experimental teachers, the control teachers did not have special training for this program and did not carry on the planned learning program in their classes.

With respect to the *Ss* themselves, in the experimental classes 50 boys and 48

girls completed all of the testing involved in this study, while in the control classes tests were completed by 40 boys and 46 girls. The average IQ for the experimental classes was 105.9 and 105.4 for the control classes as measured by the Otis Self-Administering Test of Mental Ability. As far as could be determined from school officials, no systematic method was used to assign the pupils to particular classes or to teachers (except for the two-year experimental group who were kept together the second year specifically for the orientation program.) Thus, with the *Ss* coming from comparable home backgrounds, being randomly placed in their particular class groups, and being undifferentiated with respect to average intelligence scores, the pupils in all eight classes were assumed to be originally a random sampling of a population of such pupils in their particular schools.

Statistical analyses. To test two major problems of this investigation—that of establishing significant relationships between self-ideal discrepancy and measures of manifest anxiety and observed insecurity—a three-dimensional ($2 \times 2 \times 2$) analysis of variance design was used (**10**). The factors controlled in this design were sex, the experimental-control condition, and self-acceptance. The *Ss* were divided into two groups according to their scores on the Self-Acceptance Scale—one group (the Highs) made up of those whose Self-Ideal Discrepancy scores fell above the median; the other group (the Lows) with those whose Discrepancy scores fell below the median. This median was computed using all 184 of the main group of *Ss*. In order to have equal frequencies in each cell of this design, 20 *Ss* were randomly selected from each group which represented a different combination of the factors being controlled; thus, there were 160 *Ss* used in the analysis of variance tests. The dependent variables in the analyses of variance were the scores from the Children's Manifest Anxiety Scale and the Kooker Security-Insecurity Rating Scale.

The possible effects of the orientation program on the relationship between self-ideal discrepancy and anxiety was tested by *t* tests of a difference in means between the experimental and control groups.

TABLE 1

TEST-RETEST (ONE WEEK INTERVAL) RELIABILITY CORRELATION COEFFICIENTS OF THE SEVERAL SCORES OF THE SELF-ACCEPTANCE SCALE FOR EACH OF THE TWO CLASSES STUDIED

	N	Self-Concept	Ideal-Self	Self-Ideal Discrepancy
Class I	21	.83	.69	.80
Class II	26	.93	.54	.86

RESULTS

Results of preliminary testing. Test-retest reliability of the Self-Acceptance Scale was determined by administering this test to two classes of sixth graders on two occasions, one week apart. Coefficients of correlation for the two administrations for the various sections of the scale are shown in Table 1.

Although all of these coefficients are significant beyond the one per cent level of confidence, it will be noted that those for the Ideal-Self scores are somewhat lower than the others. This can be attributed to the limited range characteristic of this particular score—that is, the pupils tended not to vary from one another in the rating of their ideal-self concepts—thus, the correlation test was particularly sensitive even to the smallest deviations from one administration to the next.

In order to test whether or not intelligence was likely to be an important factor in the self-acceptance score, IQ scores from the Otis Self-Administering Test of

Mental Ability were secured from school records for all pupils who participated in the testing program. The coefficient of correlation between the Self-Ideal Discrepancy scores and IQ scores for the 47 pupils participating in the preliminary study was −.08 indicating intelligence was not a factor.

Analyses using Self-Acceptance Scale. Tables 2 and 3 show the results of analysis of variance tests using mean scores from the Children's Manifest Anxiety Scale and the Kooker Security-Insecurity Rating Scale, respectively, when the *Ss* were divided into two groups, Highs and Lows, according to whether they fell above or below the median Self-Ideal Discrepancy score. These tables show that the differences in the means on both the anxiety and insecurity scales between the Highs and Lows were statistically significant beyond the .01 level of confidence. Furthermore, these differences were in the direction such that the group with the relatively high self-acceptance (that is, low discrepancy scores) yielded average scores indicating less manifest anxiety, as measured, and less insecurity, as rated by observers.

One qualification to these findings should be noted. In the analysis of variance tests, the interaction effects were not significant at the .05 level of confidence (the level prescribed prior to the investigation); however there was a tendency for an interaction to exist between the Self-Ideal Discrepancy scores and the experimental-control condition. (The *F* test of this interaction between the discrepancy scores and the experimental-control condition produced ratios of 3.14 and 3.29 for the Anxiety and Kooker scores, respectively, which are significant between the .10 and .05 levels of confidence.) This tendency towards interaction indicates that the discrepancy measure of self-acceptance was related to the personality variables of anxiety and insecurity differentially according to whether or not the *Ss* were in the orientation program. The implications of this possible interaction will be discussed below.

TABLE 2

MEANS AND ANALYSIS OF VARIANCE OF SCORES ON THE CHILDREN'S MANIFEST ANXIETY SCALE OF SUBJECTS DIVIDED INTO "HIGHS" AND "LOWS" ACCORDING TO THEIR SELF-IDEAL DISCREPANCY SCORES ON THE SELF-ACCEPTANCE SCALE

(N = 20 IN EACH SEX, CLASS-TYPE GROUPING)

Self-Ideal Discrepancy Score	Children's Manifest Anxiety Scale			*F*
	Boys	Girls	Total Highs & Lows	
Highs			16.78	13.40**
Exptl. Classes	12.10	18.55		
Control Classes	17.05	19.40		
Lows			12.56	
Exptl. Classes	13.05	13.25		
Control Classes	11.70	12.25		
Total (Boys & Girls)	13.48	15.86		
F	4.30*			

* Significant beyond the .05 level of confidence.
** Significant beyond the .01 level of confidence.

The results reported thus far have indicated relationship between self-acceptance, as measured, and the measures of manifest anxiety and observed insecurity. So as to obtain further information concerning the extent of this relationship, a correlational analysis of the major variables was made, and the results are reported in Table 4. The coefficients of correlation for Self-Ideal Discrepancy scores with the Anxiety scores and Kooker scores were each significant beyond the .01 level of confidence.

Effects of the orientation program. Analysis of the data indicating the effects of the orientation program on the variables meas-

TABLE 3

MEANS AND ANALYSIS OF VARIANCE OF SCORES ON THE KOOKER SECURITY-INSECURITY SCALE OF SUBJECTS DIVIDED INTO "HIGHS" AND "LOWS" ACCORDING TO THEIR SELF-IDEAL DISCREPANCY SCORES ON THE SELF-ACCEPTANCE SCALE

(N = 20 IN EACH SEX, CLASS-TYPE GROUPING)

Self-Ideal Discrepancy Score	Kooker Security-Insecurity Scale			F
	Boys	Girls	Total Highs & Lows	
Highs			25.10	13.42*
Exptl. Classes	24.90	25.65		
Control Classes	25.20	24.64		
Lows			22.81	
Exptl. Classes	22.10	21.60		
Control Classes	24.75	22.80		
Total (Boys & Girls)	24.24	23.68		
F	<1			

* Significant beyond the .01 level of confidence

TABLE 4

COEFFICIENTS OF CORRELATION BETWEEN THE SELF-IDEAL DISCREPANCY SCORE OF THE SELF-ACCEPTANCE SCALE AND THE MEASURES OF ANXIETY AND INSECURITY

	Self-Ideal Discrepancy		
	Boys N = 90	Girls N = 94	Total N = 184
Manifest Anxiety Scale	.28*	.41*	.35*
Kooker Rating Scale	.33*	.34*	.32*

* Significant beyond the .01 level of confidence

ured appears in Table 5. Separation of the experimental classes made possible a comparison of the effects between exposure to the program for one year and for two years. It will be noted that significant differences on the measures of manifest anxiety and observed insecurity appear between the classes having had the program for two years and each of the other two groups involved—those having had the program for one year and those in the control classes. These differences are in a direction which indicates that at least two years' exposure to the orientation program may serve to lower manifest anxiety and reduce observed insecurity. Significant differences between the two-year experimental group and the one-year experimental and control groups were not found with the Self-Ideal Discrepancy scores.

An interesting question is, how do the Highs and the Lows with respect to Self-Ideal Discrepancy in the two-year experimental group compare with those in the control group? This analysis is given in Table 6.

Inspection of this table shows that there is a difference in means between the High two-year experimental group and the control group in manifest anxiety which is significant beyond the .01 level of confidence. In other words, it appears that those in the two-year experimental group who retained a relatively high Self-Ideal Discrepancy appeared to indicate less manifest anxiety (as measured) than did the control *Ss* who also had a relatively high Self-Ideal Discrepancy. This lends some support to the contention that at least a two year exposure to the orientation program may allow individuals to feel more comfortable about discrepancies which they feel exist between their self-concepts and ideal-self concepts.

Inspection of Table 6 also reveals that the difference between the Highs and Lows in manifest anxiety in this two-year experimental group is not significant. In other words, it appears that the major results concerning the relationship between Self-Ideal Discrepancy scores and manifest anxiety should be qualified. High Self-Ideal Discrepancy scores may be associated

TABLE 5

MEANS AND *t* TESTS BETWEEN ONE AND TWO YEAR EXPERIMENTAL GROUPS AND BETWEEN THE TWO YEAR EXPERIMENTAL GROUP AND CONTROL GROUP ON MEASURES OF SELF-ACCEPTANCE, ANXIETY, AND INSECURITY

	N	Self-Ideal Discrepancy Score		Manifest Anxiety Scale		Kooker Rating Scale	
		$\bar{X}$	*t*	$\bar{X}$	*t*	$\bar{X}$	*t*
One-Year Exptl. Group	53	7.98		15.49		24.03	
			0.87		2.38*		2.24*
Two-Year Exptl. Group	45	7.16		11.93		22.42	
			0.82		2.30*		2.24*
Control Group	86	7.88		14.97		24.12	

* Significant beyond the .05 level of confidence.

TABLE 6

MEANS AND *t* TESTS ON THE CHILDREN'S MANIFEST ANXIETY SCALE BETWEEN THE TWO YEAR EXPERIMENTAL AND CONTROL GROUPS AND BETWEEN THE HIGHS AND LOWS IN SELF-IDEAL DISCREPANCY SCORES

	Children's Manifest Anxiety Scale			
	Highs	Lows	Difference	*t*
Two-Year Exptl. Group	11.88 (N = 16)	12.04 (N = 26)	0.16	n.s.
Control Group	18.22 (N = 40)	11.98 (N = 40)	6.24	4.00*
Difference	6.34	0.06		
t	2.92*	n.s.		

* Significant beyond the .01 level of confidence.

with high Manifest Anxiety scores only for *Ss* with insufficient self-understanding; this relationship may not apply to *Ss* experiencing an orientation program in self-understanding for at least two years. This finding also may explain the tendency for interaction effects to exist between the Self-Ideal Discrepancy scores and the experimental-control condition, particularly in the case of the Anxiety test analysis which was noted above.

DISCUSSION

The validity of the Self-Acceptance Scale. The results reported above indicated significant relationships between the Self-Ideal Discrepancy measure of self-acceptance and scores from the Children's Manifest Anxiety Scale and the Kooker Security-Insecurity Rating Scale. These findings corroborate with sixth-grade-children findings of various other investigators who used similar measures with older *Ss*. The validity of the Self-Acceptance Scale is further substantiated by the fact of its relationship to an observation measure (the Kooker scale) which was independent of the children's paper-pencil responses.

The relatively low correlation between the Kooker and the Anxiety scales (N = 184, r = .26) may indicate that these two instruments are measuring different variables, and the Self-Acceptance Scale may be tapping some of the variables distinctive to each of these other two scales as is indicated by the following coefficients of correlation (N = 184): (*a*) Correlation between Self-Ideal Discrepancy scores and Anxiety scores = .35; (*b*) Correlation between Self-Ideal Discrepancy scores and Kooker scores = .32.

The results of this investigation (particularly the analyses of the effects of the orientation program to be discussed in more detail below) raise a serious question

as to the adequacy of the self-ideal congruence concept of self-acceptance. A discrepancy between self and ideal might well mean different things to different individuals. While to one person a self-ideal discrepancy might be a threat to his self system, to another such a discrepancy might indicate that his aspirations are high and serve as a challenge to him. What seems to be important is not the discrepancy itself, but the feelings about it. Thus, as the discussion below will indicate, it isn't necessarily the high discrepancy alone which is associated with anxiety but a high discrepancy under certain conditions.

The effects of the orientation program. In the introductory discussion, it was pointed out on an a priori basis that the orientation program might be expected to have diverse effects on the behavior of children. Among the early effects of such a program may be the reduction of the anxiety which seems to accompany a high self-ideal discrepancy. Bearing this out was the tendency for interaction effects to exist, particularly in the analyses of the Anxiety scores, between the Discrepancy scores and the experimental-control condition. This indicates that the relationship between the Discrepancy scores and the measure of anxiety was not the same for the experimental classes as for the control classes.

Further evidence indicating the nature of this tendency for interaction is obtained when analysis is made of the scores of those classes in which most of the pupils had had the orientation program for a two-year period. It will be recalled from Table 6 that those with high Self-Ideal Discrepancy scores in the two-year experimental group had average Anxiety scores which were significantly lower than those in the control group. Thus support is given to the contention that at least a two year exposure to an orientation in self-understanding may allow individuals to feel more comfortable about discrepancies which they feel exist between their self-concepts and ideal-self concepts.

Indication that a period longer than one year in the orientation program may be needed for measurable changes to occur in such variables as manifest anxiety and observed insecurity was seen in the results reported in Table 5. It will be recalled that in this analysis, the two-year experimental group had average scores which indicated less manifest anxiety and less observed insecurity when comparisons were made with the one-year experimental group and the control group. Even though the differences of Self-Ideal Discrepancy scores between these groups were not statistically significant, the discussion above indicated that the feelings about a high discrepancy, where it existed, may have been changed as a result of the two-year orientation program.

Obviously, a single study is not sufficient to establish the validity of this finding. Several studies will be needed to check these results, but those from this investigation—that is, the tendency toward interaction effects noted in the analyses of the Self-Acceptance Scale and the study of the two-year experimental group—suggest that anxieties relative to high self-ideal discrepancies may be more prevalent for those with insufficient self-understanding than for those who have been trained in understanding themselves.

Summary

The purpose of this investigation was to construct a measuring instrument which could be reliably used in the investigation of purported relationships between self-acceptance and other important personality variables and in the study of the possible effects a learning program concerned with the causes of behavior might have on the participants in such a program.

The Self-Acceptance Scale, constructed to measure self-acceptance in sixth grade children, consisted of a series of descrip-

tive statements with which each child rated himself according to the extent he felt he was like the description (the self-concept) and to what extent he felt he wanted to be like the description (ideal-self). Self-acceptance was defined as the congruence (that is, relative lack of discrepancy) between self-concept and ideal-self.

The *Ss* were 184 pupils in eight sixth grade classes in different elementary schools of a medium-sized city in Iowa. Four of the eight classes had undergone a planned learning program designed to help each pupil acquire an understanding of the dynamic, variable, and complex nature of human motivation.

The other measures used in this investigation included the Children's Manifest Anxiety Scale and the Kooker Security-Insecurity Scale (ratings made by trained observers in the classrooms).

The results indicated a statistically significant relationship (beyond the .05 level) between self-acceptance, as measured by the Discrepancy scores, and the measures of manifest anxiety and observed insecurity such that those with the smaller Discrepancy scores had average scores indicating less anxiety and less insecurity. However, a tendency was noted for interaction effects to be present indicating that this relationship between self-acceptance and measures of anxiety and insecurity might have been operating differentially between the experimental and control classes.

When analyses were made of those experimental classes in which most of the pupils had had the program for two consecutive years, support was found for the contention that although Self-Ideal Discrepancy scores remained high for some of these pupils, the Anxiety scores were significantly lower than those of either the one-year experimental or the control classes. This finding suggested that the statement of relationship between self-ideal discrepancy and anxiety might have to be qualified to apply to *Ss* with insufficient self-understanding. Also, the two-year experimental group obtained average scores which indicated less manifest anxiety and less observed insecurity than did the average scores obtained by either the one-year experimental group or control group.

REFERENCES

1. Bills, R. E. A validation of changes in scores on the Index of Adjustment and Values as measures of changes of emotionality. *J. consult. Psychol.*, 1953, **17**, 135–138.
2. Brownfain, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597–606.
3. Bruce, P. A study of the self-concept in sixth grade children. Unpublished doctoral dissertation, State Univer. Iowa, 1957.
4. Castaneda, A., McCandless, B. R., & Palermo, D. The children's form of the Manifest Anxiety Scale. *Child Develpm.*, 1956, **27**, 317–326.
5. Cunningham, Ruth, Elzi, Anna, Farrell, Marie, Hall, J. A., & Roberts, Madeline. *Understanding group behavior of boys and girls.* New York: Bureau of Publ., Teachers College, Columbia Univer., 1951.
6. Jersild, A. T. *Child psychology.* (4th ed.) New York: Prentice-Hall, 1954.
7. Jourard, S. M., & Remy, R. M. Perceived parental attitudes, the self and security. *J. consult. Psychol.*, 1955, **19**, 364–366.
8. Kooker, E. An investigation of security, insecurity, achievement, and boredom in elementary school children. Unpublished doctoral dissertation, State Univer. Iowa, 1951.
9. Levitt, E. E., & Ojemann, R. H. The aims of preventive psychiatry and causality in grade school children. *J. Psychol.*, 1953, **36**, 393–400.
10. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
11. Ojemann, R. H. An integrated plan for education in human relations and mental health. *J. natl. Ass. Deans of Women*, 1953, **16**, 101–108.

12. Ojemann, R. H., Levitt, E. E., Lyle, W. H., & Whiteside, Maxine F. The effects of a causal teacher-training program and certain curricular changes on grade school children. *J. exp. Educ.*, 1955, **24**, 95–114.
13. Roberts, G. E. A study of the validity of the Index of Adjustment and Values. *J. consult. Psychol.*, 1952, **16**, 302–304.
14. Rogers, C. R. The significance of the self-regarding attitudes and perceptions. In M. L. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950.
15. Rogers, C. R. & Dymond, Rosalind (Ed.) *Psychotherapy and personality change*. Chicago: Univer. Press, 1954.
16. Secord, P. F. & Jourard, S. M. The appraisal of body-cathexis: Body-cathexis and the self. *J. cousult. Psychol.*, 1953, **17**, 343–347.

Received April 7, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 5, 1958

EFFECTS OF TRAINING IN ALTERNATIVE SOLUTIONS ON SUBSEQUENT PROBLEM SOLVING

WALTER I. ACKERMAN

Beth-El Day School, Belle Harbor, Long Island

AND

HARRY LEVIN[1]

Cornell University

An organism's adjustment to a new situation may be scrutinized for the variability or fixedness of the behavior employed. Presumably, during the early stages of solving a novel problem, behavior is characterized by its variability; as commerce with the problem increases and correct solutions are achieved, the resulting habits act to decrease subsequent variability. In a recent statement, Scott (**12**, p. 61) points out that such a sequence is characteristic even of the simplest organism:

> Such variability can be seen even in the lower organisms. If a paramecium runs into an obstacle it does not keep repeating its behavior. It backs off and approaches from a different angle, and never does exactly the same thing again. It is apparent that variability of this type is necessary for the process of adjustment, since an animal which gave fixed invariable responses could never adapt itself to a variety of changing conditions.

The problem solving model presented by the paramecium might well be envied by humans. There is ample experimental evidence that under some circumstances human problem solving is characterized by fixedness rather than variability of response. If the demands of the problem situation are similar to previously acquired solution methods, the transfer effects to novel situations should, of course, be positive. Where the demands of the problem are only superficially similar to previous successful solutions the present application of these methods is often doomed to failure. The next adaptive step would be to abandon the first method, or hypothesis, and to try another, in contrast to the persistent application of the first.

Fixedness in problem solving has been variously attributed to immediately prior problem solving experiences, to the nature of the problem at hand, and to personality dispositions of the solver (**1, 4, 5, 8**).

One major cause of lack of variability is problem solving set. Of course, should the person's set be appropriate for the problem the solution is likely to be facilitated. Under the not infrequent conditions where his set is inappropriate, the problem solver has a dual task before him, ridding himself of the maladaptive set and then applying the new solution. Recognizing Gibson's (**3**) indictment against the chaotic usage of the concept of set, we will define it for our purposes as "that manner of attacking a problem which is carried over from a previous to a succeeding problem situation."

Set precludes the variability of behavior which is essential if the habituated tack is to be cast off. This study is concerned with a way of training children

[1] This study is part of an Ed.D. thesis submitted to the Graduate School of Education, Harvard University, by the first author. The research was performed when both authors were members of the Harvard Teacher Education Research Project, supported by grants from the Fund for the Advancement of Education.

in problem solving so that in the face of new problem situations they will either not develop sets or will give them up when they do not work.

What the problem solver needs is a nonspecific approach, (sometimes called also "mode of attack" or "plan of action"), to many types of problems. One important aspect of this general approach is the assumption that problems often lend themselves to various solutions, so that if one solution does not work there is still the possibility that the problem may be solved. The solver might profitably attack the problem again, with a new orientation. How may such general problem solving behavior be inculcated? Maier (**7**) gave *Ss* specific training in reasoning and Luchins (**6**) prior to his einstellung situation instructed his *Ss*, "Don't be blind." Both investigators report that these devices aided *Ss* in overcoming habituated ways of solving problems.

This study investigates another method of inducing variability in problem solving. One group of *Ss* is taught two solutions to the same set of problems, another group one solution to these problems. Then the two groups are observed on a set of similar problems necessitating new solution methods. Next, a set is induced on a novel series of problems and the responses to situations where this set no longer works are observed. Our expectations are that the people trained in alternative solutions will solve more test problems correctly, will exhibit greater variability to these problems, and will persevere longer on problems too difficult for them to solve.

Method

The *Ss* were 48 sixth grade children, 25 boys and 23 girls, who lived in a predominantly middle class, white collar community.

Ss were divided into two groups of 24 children each, with the sexes represented as equally as possible in each group. The two groups did not differ either in intelligence or school achievement. *E* took each child individually from his classroom to a nearby experimental room. *E* told the child that he was interested in "how children solved problems." The procedure consisted mainly of training and test series on two separate types of paper and pencil problems.

Problem 1. The water jar problems. These problems are an adaptation of the water jar problems described by Luchins (**5**). *S* is required to obtain a specified amount of water using only empty jars and their total capacities as measures. The jars were drawn on cards, with a number in each jar denoting its capacity and a number in the margin indicating the amount of water to be obtained. *E* presented a single problem on a card, and *S* solved the problem on a work sheet.

The training series consisted of 10 problems, each solvable by the formulas, B − A − C, or B − 2A + C (the letters refer to the capacities of the three jars). *E* taught *S* to solve the problems by one or both of these solution methods depending on the experimental design below.

Following the training problems and a repetition of the general instructions, *S* was presented six test problems. Each of the first five called for a novel solution method; the sixth was solvable by the method taught during the training series.

S was permitted to work on each test problem until he solved it or until he decided to stop. The time spent on each problem was recorded.

Problem 2: The puzzle problems. Immediately following the completion of the above training and test series the *Ss* were given 13 puzzle problems. These are paper and pencil versions of jigsaw puzzles. *S*

was given a booklet on each page of which was a picture of a whole figure. Beneath the figure were pictured numbered fragments of various sizes and shapes. The problems call for *S* to choose those fragments which, when put together, will form the whole figure. *Ss* indicated their solutions at the bottom of the page by marking the numbers of the pieces they chose, in the order in which they used them.

The first eight of the 13 puzzles were all solvable by assembling the fragments numbered 1, 3, 5, and 7. Each of the five final problems required a different solution from the initial set-inducing series and from each other. Superficially, the test problems looked as though they were solvable by the original method. To the *S*, there was no break between the training and the test problems.

No time limit was set on these problems; a record of the time spent on each problem was kept.

Experimental manipulations. The two groups of *Ss* differed only in the treatment they received during the training series of 10 water jar problems. One group was taught the two alternative solutions to these problems (labelled A group). *E* worked the first problem by one and then the other solution method and half of the succeeding problems were worked out by one or the other method. The no alternatives group (hereafter labelled NA) were taught to solve the 10 problems by only one solution method.

Results

Stated generally, our hypotheses are (*a*) that experience in solving problems in more than one way should dispose the *S* to relinquish more quickly a maladaptive set and consequently yield more correct solutions to new problems, and (*b*) that this variability of response should generalize to problems different from those on which the dual solutions were taught.

Consider first the six test problems in the water jar series. Group A solved 111 problems correctly, whereas Group NA produced 94 correct solutions. Though the direction of difference is as perdicted, the difference is not statistically significant. More critical than the total number of correct solutions is the behavior on the first test problem. A correct answer to this problem indicates to the *S* that the solution methods of the training period may not be applicable to the ensuing problems and hence may aid him in overcoming whatever set was induced by the earlier series. Group A achieved 12 correct answers; group NA, 9. Here too the direction is interesting, but the differences are not significant.

The results are clearer when we compare the two groups in the number of other-than-training solutions they offered to the six test problems. Since there is abundant evidence that the sexes differ in their problem solving behavior, "sex" was added to "treatment" as a criterion of classification in all the analyses of variance reported.[2] As can be seen in Tables 1 and 2, Group A offered significantly more "other than training" solutions than did Group NA. We may infer then that training in alternative solutions leads to more variable problem solving behavior, though it is not clear that the increased variability yields more correct solutions.

No time limit was set on any water jar problem. The instructions directed the *Ss* to work on a problem until either a satisfactory answer was achieved or until he thought that it was no longer profitable to continue on that problem. The amount of time spent on a problem before giving up was, then, a measure of perseverance. Table 3 gives the number

[2] A correction factor for the unequal cell frequencies was applied to all of the analyses of variance.

TABLE 1

MEAN NUMBER OF OTHER-THAN-TRAINING SOLUTIONS TO TEST WATER JAR PROBLEMS

	Group A	Group NA
Male	9.33 (12)	7.15 (13)
Female	11.42 (12)	7.91 (11)

Note.—Numbers in parentheses are the numbers of *Ss* in each cell.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF OTHER-THAN-TRAINING SOLUTIONS TO TEST WATER JAR PROBLEMS

Source	*df*	Sum of Squares	Mean Square	*F*
Sex	1	2.03	2.03	1.22
Treatment	1	8.09	8.09	4.87*
Sex × Treatment	1	0.44	0.44	
Error	44	73.09	1.66	

* $P < .05$.

TABLE 3

MEAN PERSISTENCE (IN SECONDS) ON TEST WATER JAR PROBLEMS

	Group A	Group NA
Male	343.43 (7)	178.11 (9)
Female	131.41 (10)	101.84 (8)

of *Ss* in each sex and group who gave up the problems before actually achieving a correct solution and the time in seconds spent on the problems before giving up. The analysis of variance summarized in Table 4 indicates that boys persevered longer than girls and that Group A persevered longer than Group NA. The greater perseverance of boys confirms other findings about sex differences in perseverance (**10, 11**). The training in multiple possibilities of solving a problem likely taught the NA group that failure with one method did not exhaust the possibilities of success, so that they may profitably spend additional time on the problem. This finding complements Robinsons' (**9**), who reports that *Ss* who thought their chances of arriving at a correct solution to a problem were excellent spent more time on that problem than did *Ss* who were less confident.

The first part of the experiment confirms several of our expectations. There was a tendency for the group trained in alternatives to solve more problems correctly. This group offered a greater variety of solutions to the test problems, and on those problems they could not solve they worked longer before they gave up.

The puzzle problems were introduced to test whether the effects of training in alternative solutions would transfer to a

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF PERSISTENCE ON TEST WATER JAR PROBLEMS

Source	*df*	Sums of Squares	Mean Square	*F*
Sex	1	9495.53	9495.53	7.55*
Treatment	1	20777.78	20777.78	16.53**
Sex × Treatment	1	4607.02	4607.02	3.66
Error	30	37,705.80	1256.86	—

* $P < .01$.
** $P < .001$.

novel set of problems. Each of the first eight puzzles were solvable by the use of, in each case, pieces numbered 1, 3, 5, 7. The succeeding five puzzles required different pieces for their correct construction. Does training in an earlier, different type of problem influence behavior on the new test problems?

As with the water jar problems, the A group reached more correct solutions than did their controls. They solved 61 problems correctly; Group NA solved 49. On the first test problem, there were four correct solutions in the A group and none in the NA group. Though neither of these differences is statistically significant, their direction was predicted.

If the set induced by the initial eight jigsaw problems is effective, the *Ss* should continue to use the same numbered pieces in the test problems, where they are now inappropriate. Table 5 gives the mean numbers of pieces with the same numbers as those used in the training puzzles for each group. The analysis of variance for these data is summarized in Table 6. The A group evidenced greater variability by using fewer incorrect pieces than did the NA group. This finding permits us to conclude that the ability to overcome inappropriate problem solving sets is acquired through alternative training in solution methods to the same problems and that this ability is transferable to problem situations other than the one on which the alternative training was acquired.

TABLE 5

MEAN NUMBER OF SAME-AS-TRAINING PIECES (PERSISTENCE) USED IN SOLVING PUZZLES

	Group A	Group NA
Male	2.33 (12)	3.08 (13)
Female	1.83 (12)	3.18 (11)

TABLE 6

SUMMARY OF ANALYSIS OF VARIANCE OF NUMBERS OF SAME-AS-TRAINING PIECES USED IN SOLVING PUZZLES

Source	*df*	Sum of Squares	Mean of Squares	*F*
Sex	1	.04	.04	—
Treatment	1	1.10	1.10	12.37*
Sex × Treatment	1	.09	.09	—
Error	44	3.91	.089	—

* $P < .001$.

DISCUSSION

Some impressions about the water jar problems peripheral to our interests, but germane to Luchins' (5) results, occurred to us during the experiment. Luchins, it will be recalled, creates a set to use combinations of three water jars during his training problems and during the test series many *Ss* are not able to use simpler, two jar solutions. The interpretation is that set precludes *Ss* from seeing the easier solution method. In some instances in the present study it was not that the *Ss* did not see the simpler solution, but rather they did not understand that they were permitted to use two jars. Some of our *Ss* actually tried and abandoned correct two jar solutions because, as they explained it, they thought that since the initial problems involved three jars, the continued use of all three was a part of the task. It may be therefore, that in the previous water jar studies, there were people who were actually aware of the simpler solution but were not clear that it was acceptable.

The pedagogical implications of our findings are apparent. Although it involves an extrapolation from laboratory to classroom conditions, we might expect that a teacher's consistent training in alternative solutions to problems might result in efficient overcoming of set.

Put into historical perspective, the findings of this study are congruent with Woodworth's general factors theory of transfer. Apparently, what occurs when the child is given even brief training in solving problems by more than one method is that he develops a "general approach to novel problems." Colloquially, this approach is self-instruction to abandon an inefficient solution method and to try something new. One wonders whether this general problem solving skill was verbalized by our experimental *Ss* or whether it operated as an unverbalized, functional concept. Although we have no data on this point, we might expect that the verbalization of the principle would increase its efficiency.

Summary

Two groups of sixth grade children were trained to solve 10 water jar problems. One group was given a single solution to the problems; the second was taught that the problems were solvable by either of two methods. On a succeeding set of test problems which necessitated solutions different from the two training methods, the alternative group tended, though not significantly, to solve more problems correctly, offered significantly more other-than-training solutions, and persevered longer on problems they were unable to solve.

To test the transfer effects of the original training, *Ss* were given 13 jigsaw puzzles: the first eight were all solvable by a single method so that a set might occur; the final five required various solution methods. Here, also, the *Ss* who were trained in alternative methods on the earlier water jar problems evidenced greater problem solving variability by using significantly more pieces of the puzzles which were not used in the training series.

REFERENCES

1. Adamson, R. E. Functional fixedness as related to problem solving. *J. exp. Psychol.*, 1952, **44**, 288–291.
2. Birch, J. G., & Rabinowitz, H. S. The negative effect of previous experience on productive thinking. *J. exp. Psychol.*, 1951, **41**, 121–125.
3. Gibson, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.
4. Guetzkow, H. An analysis of the operation of set in problem-solving behavior. *J. Gen. Psychol.*, 1951, **45**, 219–244.
5. Luchins, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 242).
6. Luchins, A. S. Classroom experiments on mental set. *Amer. J. Psychol.*, 1946, **54**, 295–298.
7. Maier, N. R. F. An aspect of human reasoning. *British J. Psychol.*, 1933, **24**, 144–155.
8. Maier, N. R. F. The behavior mechanisms concerned with problem-solving. *Psychol. Rev.*, 1940, **47**, 43–58.
9. Robinson, Elsa E. An experimental investigation of two factors which produce stereotyped behavior in problem situations. *J. exp. Psychol.*, 1940, **27**, 394–410.
10. Ryans, D. G. The measurement of persistence. *Psychol. Bull.*, 1939, **26**, 715–739.
11. Schofield, W. Jr. An attempt to measure persistence in its relationship to academic achievement. *J. exp. Psychol.*, 1943, **33**, 440–445.
12. Scott, J. P. The genetic and environmental differentiation of behavior. In D. B. Harris (Ed.), *The concept of development*. Minneapolis: Univer. Minnesota Press, 1957.

Received April 17, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 5, 1958

THE EFFECTS OF DIFFERENT TEACHING METHODS: A METHODOLOGICAL STUDY[1]

MARVIN NACHMAN[2] AND SEYMOUR OPOCHINSKY

University of Colorado

Reviews of teaching research have consistently concluded that different teaching procedures produce little or no difference in the amount of knowledge gained by the students (**1, 2, 3, 5, 6**). This same conclusion has been reached despite the fact that experimenters have employed a wide variety of independent variables, such as lecture versus discussion classes, instructor-centered versus student-centered classes, large versus small classes, various types of TV classes, etc. These results are surprising if one considers that much of the research was instigated by the hypothesis that differences would be found. Furthermore, it appears as if most educators still assume that classroom techniques do in fact have specific effects. Why then have differences not been found?

One obvious hypothesis is that the teaching methods which have been employed are not sufficiently distinct to produce significant differences in the amount of knowledge acquired. The purpose of this paper is to examine an alternative hypothesis, namely, that the different teaching methods have, in fact, produced differential amounts of learning but that these effects have been masked in the measurement process.

Typically, in measuring the effectiveness of different teaching methods, one of the major dependent variables has been the performance of the students on the final examination. It is clear that variance on the final examination is due to many factors in addition to the specific teaching methods employed. Such things as the intellectual ability of the student, his motivations, the amount of studying he has done outside of class, various personality factors and environmental pressures, etc. will also affect his performance. In research on teaching methods, most of these other variables are not controlled (usually, the major control is to equate students for ability) and, since they undoubtedly account for a significant proportion of the variance, it is perhaps not surprising that the diverse experiments on teaching methods have so consistently concluded that there are no significant differences as a function of teaching method employed.

A basic difficulty of this conclusion stems from the fact that students prepare for examinations by studying outside of class. It is perfectly possible that two groups taught under different procedures may learn very different amounts in class, but when both then engage in significant extra amounts of study, the difference in performance becomes negligible. This is especially true since most final examinations are based, at least in part, on information to be found in the student's textbook. Thus, if both groups study the textbook equally, one might expect that they will do about equally well on the examination even though one group has learned more in class. The matter can, of course, be more complex. It may be, for example, that if Group A learns more in class than Group B, then Group B feeling itself less well prepared for the pending examination will study more than will Group A. That is, one might find that students in preparing for an examination, study until they feel they know the material to a certain degree, and if one group learns the material in class they will feel less need for addi-

[1] A portion of this paper was presented at the 1958 meeting of the APA.

[2] Now at the University of California, Riverside.

tional study than the group which has not learned the material in class.

This discussion suggests that to evaluate a particular teaching method effectively, testing of the students ought to occur immediately after the method has been employed thereby avoiding contamination by the additional variable of outside study. This procedure would be a more direct test of the influences of the teaching situation *per se* and would therefore be more likely to reveal its effects, even if these effects are not observable on a final examination.

The following experiment was designed to test this general prediction. The independent variable of class size was employed and it was hypothesized (*a*) that students in a small class would do better than students in a large class on quizzes for which they had not prepared, and (*b*) that no differences would be found between the classes when students were given an opportunity to study for an examination. The experiment was not specifically concerned with the variable of class size. Rather, it was used because it offered a convenient and simple way of testing the hypotheses and because it has so consistently been shown to have no influence on the amount learned by students when traditional measurements have been employed. In one review on the effects of class size, the author concludes that there have been "... more than 200 studies which clearly reveal that there are no consistent differences" (**4**).

Method

Subjects

Two sections of about 150 students each were enrolled in the senior author's General Psychology course. The day after the students returned from their Christmas vacation, they were informed that for the last two weeks of the semester the instructor intended to meet a small class which would be identical in all respects to the larger classes except for size. The class was arranged for Monday, Wednesday, and Friday at 8:00 A.M. and 42 of the 300 students volunteered to attend that class rather than their large one. These students were divided into two groups of 21 students and each student in one group was very closely equated with a student in the other group on the basis of the three hourly examinations they had previously taken in the course that semester. By a flip of a coin, one group of 21 was then selected for the small class and the other group remained in one of the two large classes. Thus, the *S*s for the experiment consisted of two equated groups of 21 students each, one group in a small class of 21 students and the other group in large classes of about 140 students.

Procedure

Since the experiment was not concerned with testing the unique advantages of a small class versus a large class, the teaching method used in the different classes was as nearly identical as possible. The small class met on MWF at 8:00 A.M. and the two large classes met on MWF at 10:00 and 11:00 A.M., respectively. It was arranged that the small class meet earlier in the day than the large classes so that, in the event the students communicated information to each other about examinations, the bias introduced would operate against the hypotheses. (Questioning of the students after examinations revealed that the amount of communication was just about nil.)

All three sections were conducted by the lecture method and, as much as possible, the lecturer repeated exactly the same material to the three classes. During the last 10 minutes of the second and fourth lectures, the students in all three classes were given a "pop-quiz" which specifically covered the material that was presented in that lecture and the previous lecture. The

students had typically received a few weeks notice for examinations before this time and were clearly not expecting and had not prepared for these quizzes.

In order to avoid biasing the lectures in favor of the quiz material, the quizzes were constructed by the junior author and were not seen by the lecturer until after they were administered. Both quizzes were multiple choice, the first containing 15 questions and the second 8 questions. (The second quiz was originally designed as 15 questions but since only 8 of the items were actually covered in the lectures, the rest were discarded before the quizzes were scored.)

The students took the final examination about one week after the last day of classes. The final examination consisted of 125 multiple choice items and was made up by several members of the department. It was administered at the same time to about 1,000 students, of whom 42 were the *Ss* for this experiment. Twenty-five of the items on the final examination covered the same material which had been presented while the students were in the large versus small class experimental situation. The two groups of 21 students were compared in their performance on the two quizzes and on these 25 items of the final examination.

Results

The two quizzes and the final examination were scored by assigning one point for each correct answer. For each student, a total quiz score was obtained by adding the scores of his two quizzes, and difference scores were then computed for each matched pair of students in the small and large classes. Because 13 of the 42 students were not present for one or the other of the two quizzes (seven students from the large classes and six students from the small class), the difference scores for these students were based on the one comparable quiz which both members of the matched pair had taken. The difference scores revealed that 13 of the small class students did better than their matched pairs in the large classes whereas only 3 did worse and there was no difference for 5 students. The differences analyzed with a t test for matched pairs resulted in a t of 2.57, which for 20 degrees of freedom has a probability level of less than .02.

An alternative method of analyzing the results was to correlate the two quizzes (the Pearson product-moment coefficient was .56) and to use the regression equation to predict the scores on the quizzes missed by the absent students. This resulted in two quiz scores for each student which were added. The mean of this sum was 16.02 for the small class as compared to 14.34 for the matched students in the large classes. The standard error of the difference between the means was .56 which resulted in a t of 3.00 which for seven degrees of freedom (since 13 of the scores were predicted) also has a probability level of less than .02.

On the 25 items of the final examination, 10 of the small class students did better than their matched pairs in the large classes while nine did worse, and there was no difference for two students. The mean for the small class was 18.00 as compared to 17.53 for the matched students in the large classes. The standard error of the difference between the means was .90 which resulted in a t of .53 which is not significant.

Discussion

The results confirmed the hypotheses that differential performance would be found on quizzes which specifically covered classroom material and for which the students had not prepared but that performance would be equal on final examinations for which the students had devoted a large amount of extra study. One cannot be certain, of course, about the role attributed to extra study. The spe-

cific conclusion which can be drawn from the data is that in order to test the effect of a particular classroom technique, evaluation should be done immediately after the technique is employed. Waiting until a final examination (or any other announced examination) confounds the problem by permitting many other variables to operate, one of the most obvious of which is extra study.

In most experiments on learning, when one is measuring the effects of a particular variable, e.g., massed versus distributed practice, the learning opportunities of the *Ss* are limited almost exclusively to the experimental situation. This has not been the case in experiments on teaching methods, perhaps because of the difficulties of controlling some of the extraneous variables. It may be, however, that many of the experiments on teaching methods would have led to significant differences if the evaluation procedure avoided contamination by such factors as extra study. The fact that in previous studies, the variable of class size has so repeatedly been found to produce no significant effects on amount learned, and yet yielded significant differences in the present experiment, implies that a restudy of other teaching variables using different measuring techniques might be more fruitful than it has previously been.

Although the experiment was not primarily concerned with the effects of small versus large classes, the data indicate that students in the small class learned significantly more in class than did the students in a large class. The difference of about 1.7 points on a test which had a mean score of about 15 was fairly large considering the relatively small variance of the difference scores and the fact that the small versus large class variable was limited to only two lectures per quiz. Furthermore, there was no attempt to utilize the unique advantages of a small class by permitting more questions or discussions.

What was the specific variable operating in the large and small classes which produced the difference? As Buxton (**2**) has pointed out, class size, like time, is not a variable by itself but rather is an abstraction in which other variables may be permitted to operate. In evaluating the course at the end of the semester, almost all of the students in the small class spontaneously commented that they found it easier to pay attention in the small class and that they had become more interested. Casual observations by the lecturer corroborated this. The students in the small class appeared to be much more alert in their listening as well as note-taking behavior. Rarely, if ever, were they observed in behaviors which are not uncommon in very large classes, such as talking to each other, staring out the window, reading, etc. Undoubtedly, the very proximity of the lecturer in the small class acts as an inhibitory influence on these behaviors.

It is also possible, of course, that the variable of class size had nothing to do with the obtained differences. It may be that something like a "Hawthorne-effect" was operating in which the students in the small class felt more significant and more highly motivated because a special class had been established for them. This possibility, and similar ones, such as unintentional lecturer bias, or that students perhaps learn more at 8:00 A.M. than at 10:00 or 11:00 A.M., do not reduce the significance of the major finding, namely that differences as a result of teaching technique manipulations were found on "pop quizzes" but not on a prepared-for final examination.

Summary and Conclusions

Twenty-one students in a small class were compared on examination performance with a matched group of students who were in a large class. It was hypothesized that the small class would do better on quizzes which specifically covered the

classroom material and for which the students had not prepared but that the two groups would do equally well on final examinations for which they had studied. The hypothesis was confirmed and the implications of this methodological procedure were discussed in relation to other research on the effectiveness of different teaching methods.

REFERENCES

1. Birney, R., & McKeachie, W. The teaching of psychology: A survey of research since 1942. *Psychol. Bull.*, 1955, **52,** 51–68.
2. Buxton, C. E. *College teaching, a psychologist's view.* New York: Harcourt-Brace. 1956.
3. Good, C. V. Colleges and universities—VIII. Methods of teaching. *Encyc. educ. Res.*, 1950, **42,** 273–278.
4. Kidd, J. W. The question of class size. *J. higher Educ.*, 1952, **23,** 440–444.
5. Otto, H. J., & von Borgersrode, F. Class size. *Encyc. educ. Res.*, 1950, **42,** 212–215.
6. Wolfle, D. L. The first course in psychology. *Psychol. Bull.*, 1942, **39,** 685–712.

Received April 17, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 5, 1958

CURRICULAR DIFFERENCES IN JOB INCENTIVE DIMENSIONS AMONG COLLEGE STUDENTS

A. W. BENDIG AND EUGENIA L. STILLMAN
University of Pittsburgh

In a recent study (**1**) the results of a preliminary attempt to identify the dimensions of job incentives among college students was reported. By means of a factor analysis of rankings of a homogeneous list of job incentives by college *Ss* three dimensions were isolated and tentatively identified as: (*a*) need achievement vs. fear of failure; (*b*) interest in the job itself vs. the job as an opportunity for acquiring status; and (*c*) job autonomy of supervision vs. supervisor dependency. Simple procedures were developed for computing factor "scores" for individual *Ss* from the differences between pairs of ranked incentives.

If the dimensions underlying the ranked incentives are related to choice of occupational goal among college *Ss*, we would expect that *Ss* divided into more homogeneous subgroups on the basis of their curricular courses of study would show somewhat greater agreement in their ranking of the incentives than the combined curriculum heterogeneous total group of *Ss* and that the factor scores would be capable of demonstrating significant differences among the curriculum subgroups. Curriculum subgroup differences would have the added advantage of helping to define the meaning of the job incentive dimensions.

The present research was designed to (*a*) identify, on an a priori basis, curricular subgroups of *Ss*, (*b*) combine these subgroups into major curriculum groupings by means of an empirical factor analysis of their subgroup profiles in ranking job incentives, and (*c*) test whether there were significant differences among these major curriculum groupings on the derived factor scores.

PROCEDURE

A form was prepared which listed the eight incentives used previously (**1**) and it was administered to 267 undergraduate college *Ss* (174 men and 93 women) enrolled in 10 sections of undergraduate psychology courses. The form requested the *S* to indicate his (her) name, age, sex, school, and curriculum group within the university, major subject and to write a brief description of the job he expected to accept after graduation. The *S* was then requested to *rank* (from one to eight) the following list of incentives with the incentive that would be most important to him in selecting the job previously described being ranked "one" and the least important incentive being ranked "eight." The incentives used were:

1. Opportunity to learn new skills
2. Friendly fellow workers
3. Freedom to assume responsibility
4. Good job security
5. Good prospects for advancement
6. Full insurance and retirement benefits
7. Recognition from supervisors for initiative
8. Good salary

The 267 *Ss* who completed this form were dichotomized as to sex and further divided into eight major curriculum areas:

Business Administration: 51 men.
Engineering and Mines: 42 men.
College B.A. (Humanities and Social Sciences): 35 men and 11 women.
College B.S. (Natural Sciences): 30 men and 11 women.
Pre-Education: Elementary; 26 women.
Pre-Education: Secondary; 16 men and 25 women.
Pre-Nursing: 11 women.

Nursing Education: 9 women.

This division by sex and curriculum resulted in 11 curriculum subgroups.

Two sets of "scores" were available for each *S*. The eight rankings of the incentives constituted an incentive profile for the *S* and these rankings were averaged for each of the 11 subgroups. A second set of factor "scores" had been developed in the previous study (**1**). The Factor A score was defined as the ranking of Incentive 6 minus the ranking of Incentive 1. Similarly, the Factor B score was Incentive 8 minus Incentive 2 and Factor C was Incentive 7 minus Incentive 4. Positive scores computed in this manner indicate (*a*) high need achievement, (*b*) high interest in the job itself, and (*c*) high need for autonomy from supervision. Negative factor scores presumably measure (*a*) high fear of failure, (*b*) strong attitude toward the job as a stepping-stone for advancement, and (*c*) high need for a dependency relation to the job supervisor.

Results

The average (mean) incentive profiles for each of the 11 curriculum subgroups can be found in Table 1. From the sums of ranks of the eight incentives for each subgroup the average rank-difference (rho) intercorrelation among the *Ss* constituting each subgroup was computed by the usual formula (**2,** p. 421) and these coefficients are given in the last column of Table 1. Nine of the 11 coefficients are significantly different from zero at the .01 level of confidence with the remaining two coefficients being significant at the .05 level. The average rho intercorrelation for the total group of 267 *Ss* (ignoring curriculum subgrouping) was .20 (significant at the .01 level). The median average intercorrelation of the 11 subgroups in Table 1 is .28 which can be compared to the average intercorrelation of .20 when the curriculum subgroups are combined, indicating that the division of *Ss* into curriculum subgroups did somewhat increase the homogeneity with which the *Ss* ranked

TABLE 1

MEAN RANKS OF INCENTIVES AND AVERAGE INTRAGROUP CORRELATIONS (RHO) FOR CURRICULUM SUBGROUPS

Curriculum Subgroups	Sex Group	Number of Students	Incentives								Average Intercorrelation
			1	2	3	4	5	6	7	8	
Business Administration	M	51	5.6	5.3	4.6	4.1	1.8	6.7	4.9	3.0	.38**
Engineering & Mines	M	42	4.2	4.8	4.8	4.3	2.6	7.0	5.1	3.2	.26**
College B.A.	M	35	5.3	5.1	4.1	3.8	3.2	6.3	5.3	3.0	.21**
College B.A.	F	11	4.1	3.9	2.7	4.4	4.6	6.8	4.5	5.0	.14*
College B.S.	M	30	4.6	5.4	3.1	2.8	3.5	6.5	5.9	4.2	.28**
College B.S.	F	11	2.6	5.0	3.5	3.5	3.9	7.5	5.5	4.4	.32**
Pre-Education: Elementary	F	26	3.7	4.0	3.1	2.6	5.4	6.0	6.3	4.9	.28**
Pre-Education: Secondary	M	16	4.8	4.5	4.2	3.1	3.7	6.4	4.9	4.4	.10*
Pre-Education: Secondary	F	25	3.7	4.3	2.7	2.6	5.3	7.4	5.4	4.6	.39**
Pre-Nursing	F	11	3.9	3.5	2.5	4.3	3.7	6.8	6.6	4.7	.33**
Nursing Education	F	9	2.3	6.2	2.6	4.4	3.3	7.2	5.7	4.2	.46**

* $P = .05$.
** $P = .01$.

the incentives. The median intercorrelation for the five male curriculum subgroups was .26, while the median coefficient for the six female subgroups was .32, suggesting that the female *Ss* were slightly more homogeneous in ranking the incentives than were the male *Ss*. However, the 11 coefficients were ranked and the Mann-Whitney U test (**4**, pp. 116–127) was used. The two-tailed probability was approximately .21 which indicates that the hypothesis of similar homogeneity among men and women *Ss* in ranking the incentives cannot be rejected. Two of the three smallest average intercorrelations were contributed by the college B.A. subgroups and it is suggested that in future curriculum research the two major samples of *Ss* within these subgroups, majors in the humanities and in the social sciences, be treated separately in finding incentive profiles. The intercorrelation among the Pre-Education–Secondary men might be increased by splitting these *Ss* in future samples into those majoring in physical education and *Ss* majoring in more academic high school subjects.

In order to combine the 11 curriculum subgroups into more meaningful major groupings by means of factor analysis the eight incentive mean ranks for each of the 11 curriculum subgroups were ranked and the intercorrelations among the 11 sets of ranks were computed by the rank-difference method. The mean correlation among the subgroups was .60 with individual coefficients ranging from .01 to .98. Nineteen of the 55 intercorrelations were significant at the .05 level of confidence. This matrix was factor analyzed by the usual centroid method and three orthogonal factors were extracted. The median absolute value of the residuals after extraction of the third factor was only .03, indicating the absence of a fourth factor. The original factors were rotated to simple structure and the rotated orthogonal factor loadings for the curriculum subgroups can be found in Table 2.

The 11 curriculum subgroups appeared to condense into six major groupings on the basis of their patterns of loadings on orthogonal Factors X, Y, and Z, and their factor pattern groups (I to VI) are indicated in Table 2. Factor X, which accounts for 40 per cent of the variance among the subgroups, is obviously a sex factor with the male subgroups all having high loadings on this factor and the female subgroups showing moderate or low load-

TABLE 2

FACTOR ANALYSIS OF RANK INTERCORRELATIONS AMONG CURRICULUM SUBGROUPS

Factor Pattern	Curriculum Subgroups	Sex Group	Number of Students	Factor Loadings			h^2
				X	Y	Z	
I	College B.A.	M	35	96	12	13	95
	Business Administration	M	51	94	−02	10	89
II	Engineering & Mines	M	42	74	09	58	89
III	Pre-Educ: Secondary	M	16	90	42	00	99
	College B.S.	M	30	80	54	14	95
IV	Nursing Education	F	9	37	44	71	83
	College B.S.	F	11	37	66	58	91
V	Pre-Educ: Secondary	F	25	35	89	13	93
	Pre-Educ: Elementary	F	26	33	86	13	87
	Pre-Nursing	F	11	31	76	11	69
VI	College B.A.	F	11	00	91	00	83

ings. Factor pattern Group VI is similar to Group V except for an extremely low loading on the "masculinity" Factor X which appears to justify the separation of Group VI (College B.A. women) from the three female subgroups comprising Group V. Comparing the men and women *Ss* on the mean ranks they assigned to the individual incentives indicated that the men preferred Incentives 5 (Good prospects for advancement) and 8 (Good salary), while the women reported that Incentives 1 (Opportunity to learn new skills) and 3 (Freedom to assume responsibility) would be more important in their job selection.

Factor Y (37 per cent of the intersubgroup variance) appeared to be almost the reverse of the "masculinity" Factor X with the exception that the male subgroups within Group III (Pre-Education: Secondary and College B.S.) had moderate loadings on Factor Y (median loading = .48) that were intermediate in size between the loadings of the six female subgroups (median loading = .81) and the remaining three male subgroups (median loading = .09). The factor pattern groups were trichotomized as to their loadings on Factor Y with Groups V and VI being high (median = .88), Groups III and IV being intermediate (median = .49) and Groups I and II having low loadings (median = .09). Incentives 3, 5, and 8 appeared to be related to Factor Y with subgroup preferences for Incentive 3 being positively correlated with Factor Y loadings and Incentive 5 and 8 preferences being negatively related to the loadings. Factor Y might be labelled a "tenderminded social service" vs. a "toughminded practical" dimension from an inspection of the curriculum subgroup loadings along this factor. However, "naming" of this factor is less important than is the use of the factor loadings in empirically clustering the subgroups.

Factor Z (12 per cent of intersubgroup variance) was a triplet factor with the three subgroups included in Factor Pattern Groups II and IV showing large loadings and the remaining eight curriculum subgroups showing zero loadings. Groups II and IV seem to prefer Incentive 1 more than did the remaining groups, but comparisons of the high and low Factor Z groups showed slight differences on their mean rankings of the other incentives. We might suggest that Factor Z represents an "interest in science and technology" factor, but again it should be emphasized that identification of this factor by a name was not an aim of this study.

Three factor scores had been computed for each of the 267 *Ss* in the 11 curriculum subgroups and three analyses of variance, one for each factor score, were computed to test for significant differences among the subgroups. The total variance (df = 266) of a factor score was first split into the variance between the means of the 11 subgroups (df = 10) and the residual individual differences variability among the *Ss* within the subgroups (df = 256). The 11 within-subgroup variances were tested for homogeneity by Bartlett's chi-square technique before pooling and the chi-square values gave no evidence of significant heterogeneity for any of the three factor scores. The between subgroups variance was further divided into two components: between factor pattern groups (df = 5) and the residual variability among the means of the subgroups within the factor pattern groups (df = 5). The summaries of these analyses of variance can be found in Table 3. All three factor scores discriminated among the subgroups at the .01 level of confidence. The differences among the factor pattern groups were significant at the .01 level for all three scores, while the residual variation among the subgroups comprising the pattern groups did not approach statistical

TABLE 3

ANALYSES OF VARIANCE OF CURRICULUM SUBGROUP DIFFERENCES IN FACTOR SCORES

Sources of Variation	*df*	Factor A		Factor B		Factor C	
		Mean Square	*F*	Mean Square	*F*	Mean Square	*F*
Curriculum Subgroups	10	33.98	3.80**	39.69	4.67**	31.37	3.05**
Factor Patterns	5	62.09	6.95**	73.32	8.63**	52.47	5.13**
Subgroups within Patterns	5	5.86	.66	6.05	.71	10.01	.97
Within Subgroups	256	8.94		8.50		10.29	

** Significant at the .01 level.

significance. Thus all of the variation in mean factor scores could be accounted for by the grouping of the curriculum subgroups accomplished through the factor analysis of subgroup profiles reported above.

The intercorrelations among the factor scores were computed for each of the 11 subgroups and averaged (weighted mean). The average intercorrelations ($df = 245$) among the factor scores were: A and B, $r = .16$ (significant at the .05 level); A and C, $r = -.24$ (significant at the .01 level); and B and C, $r = -.05$ (not significant). Although two of the three factor score intercorrelations are statistically significant for this large sample of *Ss*, all three are low enough to indicate that the scores are relatively independent measures of job incentive dimensions.

The mean factor scores for each of the pattern groups were computed and are given in Table 4. For Factor A, Group IV appears to be quite high in "need achievement" while Groups I and III show a high degree of "fear of failure." Groups V and VI have mean factor scores (Factor B) indicating strong interest in the job, while the other groups, particularly Group I, view the job as a means of advancing their own status and position. The average Factor C scores show Groups III and V to prefer jobs that are free of close and immediate contact with supervision and Group VI *Ss* express a need for a strong and intimate dependency relationship with their supervisors.

TABLE 4

MEAN FACTOR SCORES OF THE CURRICULUM FACTOR PATTERN GROUPS

Factor Pattern	*N*	Factor A	Factor B	Factor C
I	86	1.06	−2.17	1.06
II	42	2.69	−1.55	.76
III	46	1.87	− .80	2.63
IV	20	4.90	−1.25	1.60
V	62	2.95	.73	3.11
VI	11	2.73	1.09	.09

DISCUSSION

From one point of view the present study had two questions: (*a*) whether the somewhat crude factor "scores" developed in the previous study (**1**) were valid in discriminating differences between subgroups of *Ss* who had chosen different college curricula, and (*b*) whether a factor analytic grouping of the curriculum subgroups would account for a sizable amount of the intersubgroup variability in factor scores and provide information about the dimensions of job incentives for college *Ss*. Table 1 indicates that several mistakes were originally made in defining the (assumed) homogeneous curriculum subgroups. Both the men and women

college B.A. subgroups should have been further split to achieve more homogeneity; probably by separating *Ss* majoring in the Humanities from those majoring in the Social Sciences. Similarly, the men in the Pre-Education subgroup should be divided into more similar subgroups. However, these errors in the a priori subgrouping of the *Ss* could not be corrected after the Table 1 results were known without leaving the study open to the charge of data manipulation. These subgrouping errors appear to have had little effect on the answers to the original two questions.

The 11 subgroups condensed nicely into six broader groupings on the basis of the factor analysis reported in Table 2. Although we labelled the three obtained Factors X, Y, and Z, as "masculinity," "tenderminded social service vs. toughminded practicality," and "interest in science and technology," it must again be emphasized that we are not convinced that there are adequate appellations and that "naming" of the factors was unimportant for our purpose. We were interested solely in the use of factor analysis here to allow us to combine curriculum subgroups into a more parsimonious set of major groupings.

The answers to the two major research questions are found in Table 3. The factor scores showed differences among the 11 curriculum subgroups that were highly significant. Parenthetically it might be noted that the data given in Table 3 allows the computation of epsilon correlation coefficients to provide rough indices of the "validity" with which the Factor A, B, and C scores discriminated differences among the subgroups. Epsilon is a curvilinear correlation method similar to the eta correlation coefficient included in most statistics texts and is based upon the ratio of the variance of the means of the subgroups to the variance within the subgroups (**3,** pp. 319–324). These validity coefficients were .31, .35, and .27. The average intercorrelations among the factor scores were also low enough to estimate that each score provided some independent information about subgroup differences. Table 3 also shows that practically all of the subgroup differences in scores were accounted for by the differences among the factor pattern groups. Another way of saying this is that the curriculum subgroups included within the six larger groupings were quite homogeneous in factor scores. Curriculum differences account for approximately 10, 12, and 7 per cent of the inter-subject variability in factor scores with the six factor pattern groups accounting for practically all of this curriculum-related variance.

It seems fair to conclude from this preliminary study that (*a*) the method of factor analysis provides a technique by which the dimensions of job incentives can be isolated and relatively independent factor scores can be estimated for individual *Ss*, and (*b*) these factor scores are significantly related to college *Ss*' choices of academic curricula. However, further expansion and delineation of the factors underlying job incentives is needed with a consequent development of additional factor scores that can be estimated more reliably.

Summary

A list of eight job incentives was ranked by 267 college *Ss* and the *Ss* were divided into 11 subgroups on the basis of sex and college curriculum variables. A factor analysis of the subgroup incentive profiles isolated three factors (tentatively labelled "masculinity," "tenderminded social vs. toughminded practicality," and "interest in science and technology") and condensed the 11 subgroups into six major curriculum groupings on the basis of the subgroup patterns of factor loadings. Scores on three factors isolated in a previous factor analysis of the incentives were computed for individual *Ss* and significant differences

(.01 level) were found among the six factor pattern groups for all three factor scores.

REFERENCES

1. Bendig, A. W., & Stillman, Eugenia L. Dimensions of job incentives among college students. *J. appl. Psychol.*, in Press.
2. Lyerly, S. B. The average Spearman rank correlation coefficient. *Psychometrika*, 1952, **4**, 421–428.
3. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.
4. Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Received April 12, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 5, 1958

RELATIONSHIP OF INTELLIGENCE AND SOCIAL POWER TO THE INTERPERSONAL BEHAVIOR OF CHILDREN[1]

ALVIN ZANDER AND ELMER VAN EGMOND

Research Center for Group Dynamics, University of Michigan

There are contradictory beliefs about the behavior of highly intelligent children in school, and particularly about their participation in problem-solving discussions. These children are described as both influential and impotent, tolerant and impatient, supporting and rejecting, eager and bored. Little information exists which can help us to separate fact from fancy among these assertions, nor do we know much more about the ways in which intelligent persons differ from less intelligent ones in these respects.

There is a reason for this lack of information. Intelligence as a concept is primarily intended to describe an ability to deal with cognitive problems. There are few elements in definitions of intelligence which suggest that variations in intellectual ability are associated with variations in face to face behavior. Hence, intelligence has seldom been used as an independent variable in studies of social behavior.

In a decision making group, however, a person with high intelligence may perhaps offer wiser observations than one with low intelligence. Because of his greater usefulness, we can anticipate that a brighter child would be more influential than a less intelligent one. Because he is more expert, he should have greater social power, that is, a greater ability to influence others. On the basis of assumptions like these, it is apparent that intelligence can be a cause for particular types of interpersonal behavior. Some children, regardless of their degree of intelligence are able to exercise strong influence on their peers, while others are consistently ignored by classmates.

How then does a person's intelligence affect the way he acts toward others when his group must reach a decision? Does he behave differently when he is used to having his ideas accepted (has high social power) than when he is used to being ignored? These are primary interests in this study.

Boys and girls are expected and required to behave differently in social settings. Boys, for example, are more often pressed than are girls to be concerned with achievement and influence. Because prescriptions for the two sexes differ, it is probable that the meaning and effects of intelligence or social power differ for boys and girls. A secondary purpose of this study is to examine the impact of intelligence and social power on the interpersonal behavior of boys compared to girls.

Toward these purposes, measures were made of the intelligence and social power of all children in a number of classrooms. These persons were then put in standardized, small, problem-solving groups. Their participation was observed in terms of precoded categories to see how those with different degrees of power and intelligence differed in their behavior. Data were also obtained from teachers and classmates concerning characteristics of these children in regular classrooms.

Major Issues

In the small discussion groups, we assumed children would behave in the ways

[1] The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare.

they typically do in their classes to try to win acceptance for their ideas.

It was expected that intelligence should make a difference in the actions initiated by them, as already mentioned, because intelligent persons have more resources to offer in a problem-solving discussion than do less intelligent ones. Also, they might have more confidence in their own proposals, stemming from the ready acceptance of them in the past. Highly intelligent pupils, therefore, were expected to make more efforts to influence others, to have their ideas accepted more often than less intelligent pupils, and to behave in ways which could be taken as typical of persons with confidence in themselves.

Children with greater social power were expected to make more efforts to influence others, to be more successful in doing so, and to behave in ways which have been shown in other studies to be typical of persons with greater power (**2, 3, 4, 5, 6**).

How do teachers characterize children who have different degrees of intelligence and social power? To determine this, teachers were asked to rate the behavior of *Ss* in categories roughly similar to those used in observing the behavior of the children in the small groups. What aspects of the teachers' opinions, based on day by day experience with these children, support or contradict the behavior shown in the standardized group situations? Were teachers able to differentiate between the behavior of one type of person and that of another?

Finally, it is useful to know how classmates characterized the children to whom they ascribed high social power as compared to those to whom they attributed little power. The *Ss* were rated by their peers concerning their ability in schoolwork, their attractiveness, and their ability to coerce or threaten others. These personal qualities were considered to be separate sources of social power as suggested by French and Raven (**2**). It was anticipated that persons with greater social power, in contrast to those with less, would have these qualities ascribed to them more often by classmates. Do highly intelligent persons differ from the less intelligent in these respects?

Method

Data used in this research were originally collected by colleagues for a different but related purpose.[a] In the original investigation, measures were obtained concerning all children in 16 second grade and 16 fifth grade classrooms, representing all socioeconomic levels in a medium-sized city. Children were selected in that study for a field experiment concerned with the creation of changes in the social position and behavior of group members. The measures were made in order to establish a baseline, so that the amount of change in a participant's behavior could be determined. The data employed in the present investigation are from these pre-experimental measures.

From the original population of 638 children on whom measurements were available, *Ss* were chosen for this study who had degrees of intelligence and power in required combinations. Pupils designated as high in intelligence were those in the upper 33 per cent of their class and those designated as low in intelligence were in the lower 33 per cent. Children designated as high in power were in the upper 50 per cent of their class and those designated as low in power were in the lower 50 per cent of their class on this measure. The sample included 226 boys

[a] The collection of the original data was supported by Grant-in-Aid (M-919) from the National Institute of Mental Health, National Institutes of Health, US Public Health Service. Ronald Lippitt was project director. We are grateful for his kindness in furnishing the data for our use.

and 192 girls, 230 second grade children and 188 fifth graders.

The type of data available for each child and the source of each are as follows:

Intelligence. Intelligence scores were obtained from school records. They were based on the results of the Kuhlman-Anderson test, administered in a group form.

Ratings by classmates. Every class member rated every child in his class on four characteristics: social power (who can most often get you to do things for him?), attractiveness, ability in school work, and ability to threaten others. To obtain these ratings, photographs of every child in the classroom were printed on sheets with a four-point rating scale next to each picture.

Observed behavior in groups. The members of a class were divided into four smaller groups on a random basis and each group was sent to a corner of the room where they worked on assigned problems. A trained observer was stationed in each corner who recorded on a precoded observation schedule the quantity and quality of behavior initiated by each child.[3] Four problems were assigned to the groups, each of which required a group decision as a first step in progress toward completion of the task. In one problem, for example, the group built a large tinker-toy and in another arrived at a decision as to how many beans were in a bottle. At the end of each task new groups were formed for the next problem, thus providing maximum opportunity for each child to interact with every other child in the room. No chairmen were designated for these discussions.

The observed behavior was coded into the following categories:

1. Influence attempts—efforts to influence others regardless of the manner employed.
2. Successful influence attempts—efforts in which compliance was obtained from others.
3. Unsuccessful influence attempts—efforts in which compliance was not obtained from others.
4. Demanding influence attempts—comments made in an ordering or directing manner, implying autonomy for the actor.
5. Suggestions—comments indicating weak proposals to, or requests of, others.
6. Valuing of others—behavior indicating recognition of either high or low resourcefulness of another person in an area of knowledge or skill.
7. Positively valuing others' behavior—comments indicating recognition of high value in others' behavior.
8. Negatively valuing others' behavior—comments indicating recognition of low value in others' behavior.
9. Affect-laden behavior—behavior in which overt friendliness or unfriendliness is observed which is not a direct attempt to influence another.
10. Aggressive behavior—acts of aggression which are either inflicted or threatened.
11. Mean weighted directness in style—ratio of frequency of forceful to nonforceful forms of behavior toward others.

Perceptions by teachers. Every child was rated by his teacher on seven characteristics descriptive of the typical social behavior of the person in his schoolroom. The teacher was asked to indicate on a five-point scale the extent to which each child showed the behavior under consideration (from "hardly ever" to "most often"). The qualities rated are shown in Table 6.

Teachers did not know how much social power classmates had attributed to the *Ss* at the time they made their own ratings. Information concerning the intelligence of these children was available to teachers in school records.

Results

Characteristics attributed to the *Ss* by their peers, for boys and girls with different degrees of power, are first discussed. Results are then presented for (*a*) the observed behavior of *Ss* in the small groups and (*b*) the perceptions of *Ss*' daily behavior by teachers. Finally, the results are summarized and interpreted.

For the sake of brevity the data in tables are usually limited to statistically significant findings. The omission of results for a specific category of behavior indicates that no significant differences were obtained for it.

At the outset it should be noted that

[3] The observers had satisfactory reliability among them. Information concerning reliability of the observations will be available in publications concerning the project from which these data were borrowed.

social power is not highly correlated with intelligence. The correlation for boys was .20 and for girls .28. The low relationships between intelligence and power indicates that they may vary quite independently. Since both of these correlations are significant at the .01 level of probability, however, it is also clear that brighter children tend to have more power than less intelligent ones.

Characteristics Attributed to Subjects by Classmates

Both boys and girls who were attributed high social power were more attractive to classmates than those low in power, regardless of their intelligence. These results may be seen in Table 1. Girls with greater power were rated as more able in schoolwork than girls low in power irrespective of their intelligence, while boys were described as more able in schoolwork to a significant degree only among those high in power and intelligence. Boys with higher power were described as more threatening. This was not true for girls.

Boys and girls with higher intelligence were seen by classmates as significantly more able in schoolwork than less intelligent persons (M 2.24 and 1.76, p of diff. = .01). Girls high in intelligence were rated as more attractive than those low in intelligence (M 2.40 and 2.01, p of diff. = .005). Ability to threaten was not related significantly to intelligence.

Behavior of Boys in Small Groups

We consider the types of behavior boys used in the problem solving discussions.

Effects of Variations in Power

Among highly intelligent boys, those high in social power were not significantly different in any category of observed behavior from those low in power.

Among less intelligent boys, those high in power, compared to those low in power, revealed a vigorous, inconsistent, and com-

TABLE 1

Means of Characteristics Attributed by Classmates to Children with Varied Social Power

Children high in intelligence	Boys			Girls		
	Social Power		t	Social Power		t
	High M	Low M		High M	Low M	
Attractive	2.13	1.52	6.42**	2.09	1.50	7.66**
Able in school work	2.19	1.43	6.33**	2.14	1.58	7.18**
Threatening	3.02	2.76	3.29**	2.89	2.86	.39
N	58	38		73	42	
Children low in intelligence						
Attractive	2.21	1.59	6.59**	2.39	1.65	11.69**
Able in school work	2.44	1.73	1.15	2.55	1.76	16.28**
Threatening	2.91	2.78	1.86*	2.98	2.90	.90
N	46	84		26	51	

* $p = .05$.
** $p = .001$.

petitive pattern of behavior. The quantitative results may be seen in Table 2.

Effects of Variations in Intelligence

Among boys with high power, the behavior of the more intelligent was in no way significantly different from that shown by the less intelligent.

Among boys with low power, those with greater intelligence, compared to those with less, were active, effective, and supportive of others. These data are summarized in Table 3.

Behavior of Girls in Small Groups

Effects of Variations in Power

Among girls with high intelligence, those with high power were not observed to behave differently from those with low power.

Among girls who were less intelligent, those with high social power were only a little different from ones low in power. Those with more power were more often successful in their influence attempts (M 8.81 and 5.61, p of diff. = .01) and were more positive in their remarks about others' behavior (M 1.28 and .58, p of diff. = .01).

Effects of Variations in Intelligence

Regardless of their social power, variations in intelligence were not associated to a significant degree with differences in girls' observed behavior.

Behavior of Boys and Girls Directly Compared

A direct comparison of the behavior used by boys and girls, regardless of their power or intelligence indicates the actions most typical of each sex. The following behaviors were observed among boys significantly more often than among girls: attempts to influence, successful influences, unsuccessful influences, aggression, and demands. Girls did not display any type of behavior significantly more often than boys. It is evident that boys were considerably more active and demanding in their groups than were girls.

TABLE 2

Behavior of Less Intelligent Boys Varying in Social Power

	Social Power		t
	High M	Low M	
Total influence attempts	24.34	16.00	3.18**
Freq. successful influence attempts	13.87	7.96	3.16**
Freq. unsuccessful influence attempts	10.48	8.04	1.98*
Freq. demanding influence attempts	5.33	3.46	2.10*
Freq. suggestions	18.90	12.30	3.13**
Total valuing of others' behavior	2.48	1.73	2.03*
N	46	84	

* p = .05.
** p = .001.

TABLE 3

Group Behavior of Boys with Low Social Power Varying in Intelligence

	Intelligence		t
	High M	Low M	
Total influence attempts	21.63	16.00	2.02*
Freq. successful influence attempts	11.97	7.96	2.06*
Freq. suggestions	16.86	12.30	2.04*
Freq. positively valuing others' behavior	1.37	.89	2.58**
N	38	84	

* p = .05.
** p = .01.

TABLE 4

BEHAVIOR OF HIGHLY INTELLIGENT BOYS VERSUS GIRLS VARYING IN SOCIAL POWER

	Social Power			Social Power		
	Low Boys *M*	Low Girls *M*	*t*	High Boys *M*	High Girls *M*	*t*
Total influence attempts	21.63	13.57	2.30*	25.55	16.69	3.08***
Freq. successful influence attempts	11.97	7.21	1.80*	15.76	9.79	2.62***
Freq. unsuccessful influence attempts	9.66	6.36	2.39**	9.78	6.90	1.79*
Freq. suggestions	6.87	10.04	3.13***	19.95	13.80	2.75***
Total valuing of others' behavior	2.16	1.67	.97	2.29	2.09	.48
Freq. demanding influence attempts	4.63	3.14	.28	5.17	2.61	2.64***
Freq. aggressive behavior	2.97	1.50	1.89*	2.79	1.54	2.36**
Total affect-laden remarks	1.95	2.02	.15	2.46	1.63	3.92***
M weighted demandingness in behav.	21.15	20.95	.31	22.64	21.25	1.50
N	38	42		58	73	

* $p = .05$.
** $p = .01$.
*** $p = .001$.

When members of both sexes were high in intelligence but low in power, boys were more active and aggressive than were girls. These data are shown in Table 4.

Where members of both sexes were high in both intelligence and power, the patterns of behavior were similar to those just described. In addition, high-high boys were demanding in their comments and used more affect-laden types of behavior than high-high girls (see Table 4).

In contrast to the girls, then, highly intelligent boys were likely to be active in their groups regardless of their social power and likely to be aggressively insistent in stating their opinions when high in both power and intelligence.

Among the less intelligent children the boys again appeared to be more involved in their groups than the girls if they were high in power. A comparison of the behavior of boys and girls is presented in Table 5. Boys with low intelligence and low power were very little different from girls with low intelligence and power.

TEACHERS' PERCEPTIONS OF BOYS

We turn to an examination of the qualities teachers attributed to these children.

Effects of Variations in Power

Among highly intelligent boys, the teachers made clear distinctions on every category between boys who were high in power and those who were low in power. These results are shown in Table 6.

Among less intelligent boys, the teachers made similar distinctions in the characteristics they attributed to boys high in power and those low in power. These re-

TABLE 5

BEHAVIOR OF LESS INTELLIGENT BOYS VERSUS GIRLS AS RELATED TO SOCIAL POWER

	Social Power			Social Power		
	High Boys *M*	High Girls *M*	*t*	Low Boys *M*	Low Girls *M*	*t*
Total influence attempts	24.34	15.96	2.44**	16.00	13.49	1.09
Freq. successful influence attempts	13.87	8.81	1.99*	7.96	5.61	1.59
Freq. unsuccessful influence attempts	10.48	7.15	1.90*	8.04	7.88	.12
Freq. suggestions	8.89	13.23	2.07***	12.31	9.75	.84
Freq. positively valuing others' behavior	1.19	.61	2.04*	.89	.58	1.63*
Total valuing others' behavior	2.48	1.54	1.96*	1.76	1.41	.99
M. weighted demandingness	21.98	19.70	.30	21.95	20.09	2.86***
Freq. aggressive behavior	3.24	2.00	1.40	3.13	2.20	1.50
N	46	26		84	51	

* $p = .05$.
** $p = .01$.
*** $p = .001$.

sults are listed in Table 6. In the eyes of teachers, boys with greater power are strikingly different in their social behavior from those with less power.

Effects of Variations in Intelligence

Teachers made no distinctions to a significant degree between boys high in intelligence and those low in intelligence, regardless of their social power.

TEACHERS' PERCEPTIONS OF GIRLS

Effects of Variation in Power

Among girls high in intelligence, those with high social power were seen to be different from girls low in power as shown in Table 7.

Among girls low in intelligence those with high social power were perceived by teachers to be different from those with low social power in only two respects. Girls with higher power were seen as more often successful in influencing others (*M* 4.82 and 3.77, *p* of diff. = .005) and more friendly (*M* 4.30 and 3.46, *p* of diff. = .005).

Effects of Variations in Intelligence

Among girls with greater social power, teachers saw girls with high intelligence as being little different in their classrooms from those with low intelligence: the girls were viewed as making more attempts to influence classmates (*M* 4.26 and 3.80, *p* of diff. = .02) and as more successful in these attempts (*M* 3.77 and 3.19, *p* of diff. = .005).

Among girls with little social power, highly intelligent girls compared to less intelligent ones were seen by teachers as making more frequent attempts to exercise influence in the class (*M* 4.29 and 3.76, *p* of diff. = .02), as more successful in these

TABLE 6

CHARACTERISTICS TEACHERS ATTRIBUTE TO BOYS WITH DIFFERENT DEGREES OF SOCIAL POWER

	Boys high in intelligence			Boys low in intelligence		
	Social Power			Social Power		
	High *M*	Low *M*	*t*	High *M*	Low *M*	*t*
Freq. influence attempts	4.05	3.70	1.50	4.10	3.52	2.40**
Amt. success as influencer	4.10	3.24	3.69***	4.46	3.50	4.38***
Friendliness	3.95	3.41	2.43**	4.27	3.41	3.63***
Depends on teacher	3.57	4.62	3.84***	3.76	4.32	3.08***
Depends on peers	3.68	4.45	3.00***	3.56	4.04	2.23*
Self-centeredness	3.58	4.26	2.93***	4.03	3.68	1.60
Degree of forcefulness	3.68	4.44	3.00***	3.56	4.04	2.23*
N	58	38		46	84	

* $p = .05$.
** $p = .01$.
*** $p = .001$.

efforts (*M* 4.82 and 3.78, *p* of diff. = .005), and as more friendly (*M* 4.29 and 3.62, *p* of diff. = .005).

TEACHERS' PERCEPTIONS OF BOYS AND GIRLS COMPARED

The perceptions teachers had of boys and girls differed to a significant degree only where the members of both sexes were high in power while low in intelligence. Here, boys were seen as more active in making attempts to influence classmates and more self-centered in doing so than were girls.

DIFFERENCES DUE TO AGES OF SUBJECTS

The results on all dimensions for second graders were compared with those of fifth graders. There was no acceptable evidence that differences in ratings of *Ss* by peers or teachers, or in their observed behavior in the small groups, were associated with differences in age.

TABLE 7

CHARACTERISTICS TEACHERS ATTRIBUTE TO HIGHLY INTELLIGENT GIRLS WITH DIFFERENT DEGREES OF SOCIAL POWER

	Social Power		*t*
	High *M*	Low *M*	
Amt. success as influencer	3.78	3.19	3.39**
Friendliness	3.62	3.07	2.75**
Depends on teacher	4.00	4.38	1.80*
Self-centeredness	3.93	4.63	3.16**
N	73	42	

* $p = .05$.
** $p = .001$.

SUMMARY AND DISCUSSION OF RESULTS

Characteristics Attributed by Peers

Boys and girls were similar in that those with greater power, compared to those with less, were better liked by their peers.

Attractiveness then was an important basis of power for both sexes.

Boys and girls were different in that boys with greater power were seen by classmates as more threatening, and girls with greater power were rated as more expert "in the things you do in school."

These results suggest that among these children the two sexes won social power in part on the basis of different attributes, boys earned it by being threatening and girls earned it by being skilled in the things required of a school child.

Observed Behavior in Groups

The behavior of boys may be briefly summarized by noting that those who were low in social power and low in intelligence (low-lows) were strikingly different from boys with all other combinations of intelligence and power. The significant differences we have seen in the behaviors of boys stem primarily from the fact that the low-lows were passive persons in their groups, more than boys who were high in either power or intelligence. The low-lows significantly less often tried to influence others, were less successful in doing so, were less demanding in manner, less often evaluating in respect to others' contributions, and less often suggestors of tentative proposals. Boys with various combinations of power and intelligence other than low-low were in no way different from one another to a statistically significant degree in their observed behavior, that is, high-highs were not different from high-lows or from low-highs. It is worth special note that highly intelligent boys whose social power was low behaved in no way different from those whose social power was high.

Because the low-lows were different from boys with more power or intelligence in ways that were quite comparable, and because the high-highs were in no way different from the high-lows or low-highs; it appears that the possession of greater intelligence may be the same as the possession of greater power insofar as the effect upon behavior in these groups is concerned.

This similarity is due, we believe, to the fact that the contributions made by highly intelligent boys in a problem solving discussion represent resources which are valuable to the group. The possession of these resources, we assume, provided power for the owners to influence those who valued them. It is highly probable, although it cannot be demonstrated with the data available here, that boys with high intelligence were treated by group members as though they were persons with high power. As a consequence, boys with high intelligence became aware of the value of their ideas and of the influence they were having on the discussion. Although they had little power accorded to them previously by their classmates, boys with high intelligence apparently behaved in the groups like those who had come to the group with social power already attributed to them by their peers.

It is noteworthy that boys low in intelligence yet high in social power tended to be more inconsistent than low-lows in that they made both demands and weak proposals, and praised as well as criticized others, whereas boys high in intelligence but low in power more consistently proposed ideas and supported others' behavior than did the low-lows. The possibility is suggested by these findings that greater power among boys generated more inconsistent and coercive group behavior while greater intelligence evoked a consistency and readiness to be considerate in relations with others. This conjecture is supported by the findings that peers characterized boys with greater power as threatening but did not so describe boys with greater intelligence.

Girls who were low-lows were less suc-

cessful in influencing their groups and less often made positive remarks than girls high in power but low in intelligence. This indicates that low-low girls, like the low-low boys, were passive in their groups' discussions. Low-low girls were not different in any respect from girls low in power and high in intelligence, which suggests that high intelligence was not similar to power among girls, as was earlier noted for boys.

In most combinations of intelligence and power, boys were significantly more active in seeking to influence others, more often successful, more often unsuccessful, and more likely to evaluate the comments made by others than were girls. Low-low boys were very much like girls and most like low-low girls, differing from them in only two respects: they were significantly more likely to be demanding and to be positive in commenting upon the contributions made by others.

In sum, the possession of either power or intelligence by boys appeared to stimulate vigorous and successful participation in their groups' work while the possession of low power and low intelligence together generated passivity in boys. The amount of intelligence or power a young person possesses affected a boy's behavior more than it did a girl's behavior in these problem solving discussions.

To explain the consequences of power and intelligence for boys and girls we assume that these two attributes have a differential significance for the sexes in allowing them to conform to the expectations society puts upon them. Barry, Bacon, and Child (**1**) have reported that boys are expected to be self-reliant and to strive for achievements, while girls are urged to be nurturant, obedient, and responsible in almost all societies, including ours.

We assume that, in their group behavior, the boys and girls were attempting to conform to these prescriptions for their sexes. Boys who were high in either intelligence or social power more clearly fulfilled their sex roles than those lacking in both of these. Boys who were low in both intelligence and power least often showed the behavior required of their sex. The low degree of their intelligence and power apparently made them unable to perform in ways typically expected of them. Thus, either social power or intelligence was necessary for boys if they were to act as boys are expected to act. There is some indication that social power was more important than intelligence in this respect.

Girls who were high in power and intelligence were little different from those who were low in either of these qualities because, we believe, high social power and intelligence were not needed in order to be the nurturant, obedient, or responsible persons required by society. Girls could fulfill these expectations regardless of the amount of power or intelligence they possessed.

Perceptions by Teachers

The children who were accorded higher social power by their classmates were viewed by teachers as most influential, since the teachers rated boys and girls with higher power as more successful in influencing others than those low in social power.

On almost every category, teachers distinguished between boys low in power and those high in power, regardless of the boys' intellectual abilities. It is striking in this connection that teachers saw those with greater power as less forceful and more friendly than those with less power, whereas boys with greater power (and low intelligence) were demanding and less friendly than those with less power when observed in the small groups. Clearly teachers did not perceive the boys with

greater social power as threatening persons in the way that classmates saw them.

In their considerations of girls, apparently those who were high in either intelligence or power were seen as more effective members of their classes than those low in these respects.

Why did the teachers characterize boys who had greater power (but low intelligence) as more considerate of others than was observed among them in the small groups? The most likely explanation is that their behavior actually was different in their classrooms from what it was in the discussions. In a class the teacher is in charge so that nondemanding, friendly behavior is required and becomes the standard way for interacting with peers during school. It is also probable that teachers approved of achievement oriented and influential efforts when they were used by boys since such behavior is in accord with the demands that teachers, among other adults, place upon young males. Thus, teachers attributed positive characteristics to the boys whom they viewed as most influential in their classrooms.

In the small groups no one was in charge. Hence, the boys with greater power, but less intelligence, were free to insist upon having their opinions accepted and were free to use coercion when necessary in order to be influential.

Teachers did not perceive differences in classroom behavior between boys who were high in intelligence and those who were low in intelligence although differences between these two groups were evident in the problem solving discussions. It appears that variations in intelligence do not generate variations in social behavior in a classroom, of the kind that teachers were asked to report.

We assume that girls were less active in their classes and therefore less visible to the teachers than the boys. Thus, teachers made fewer distinctions in characterizing the behavior among girls with different amounts of power than they did among boys. Girls with higher power were seen by teachers as most friendly to classmates, which suggests that through this friendliness they won influence in the classroom. Girls with higher intelligence were also seen by teachers as being more friendly and as making more attempts to influence others than girls with low intelligence. It is probable that the girls with high intelligence did use acceptable forms of social influence in their schoolroom relations since they were rated as attractive by their classmates, more so than those with low intelligence. This acceptance by their peers apparently gave them confidence to exercise their influence freely and teachers noted this in characterizing highly intelligent girls.

In summary, teachers perceived the behavior of girls as pretty much alike regardless of their power or intelligence. They made no significant distinctions among boys with different degrees of intelligence, but they saw many distinctions in the behavior of boys who differed in social power. Assuming that the teachers' perceptions are accurate, it is evident that a boy's social power determines his behavior in the classroom more than his intelligence does, whereas differences in the power or intelligence of a girl has little effect upon the behavior she employs in the schoolroom.

REFERENCES

1. Barry, H., Bacon, Margaret, & Child, I. A cross-cultural survey of some sex differences in socialization. *J. abnorm. soc. Psychol.*, 1957, **55**, 327–332.
2. French, J. R. P., Jr., & Raven, B. The bases of social power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Institute for Social Research, in press.
3. Hurwitz, J., Zander, A., & Hymovitch, B. Some effects of power on the relations

among group members. In D. Cartwright and A. Zander (Ed.), *Group dynamics research and theory*. Evanston, Ill.; Row, Peterson, 1953, 483–492.
4. Lippitt, R., Polansky, N., Redl, F., & Rosen, S. The dynamics of power. *Hum. Relat.*, 1952, **5**, 37–64.
5. Zander, A. & Cohen, A. R. Attributed social power and group acceptance. *J. abnorm. soc. Psychol.*, 1955, **51**, 490–492.
6. Zander, A., Cohen, A. R., & Stotland, E. *Role relations in the mental health professions*. Ann Arbor, Mich.: Institute for Social Research, 1957.

Received June 4, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 5, 1958

AN APPROACH TO THE MEASUREMENT OF PSYCHOLOGICAL CHARACTERISTICS OF COLLEGE ENVIRONMENTS[1]

C. ROBERT PACE AND GEORGE G. STERN
Syracuse University

The present article considers the idea that college cultures may be seen as a complex of environmental press which, in turn, may be related to a corresponding complex of personal needs. In the psychological literature, one is indebted to Henry Murray (**5**) for the dual concept of personal needs and environmental press. In the broadest sense, the term "need" refers to denotable characteristics of individuals, including drives, motives, goals, etc. The term "press" can similarly be regarded as a general label for stimulus, treatment, or process variables. Murray's concept of needs has provided a starting point for the construction of various objective measures of personality (**2, 3, 4**). No parallel development in the objective measurement of environmental press, however, has previously been attempted. College students differ. College environments also differ. The concept of press offers a way of viewing the environment which is comparable analytically and synthetically to the more familiar ways of dealing with the individual.

DEVELOPMENT OF THE COLLEGE CHARACTERISTICS INDEX

Using Murray's classification of needs as a model, Stern has constructed several experimental editions of a needs inventory, called the Activities Index. In its current form the Activities Index consists of 300 statements of commonplace, socially acceptable activities to which responses of "like-dislike" are given. There are 30 scales of 10 items each, corresponding to 30 needs in Murray's taxonomy. Some scales can be scored positively or negatively, as for example conjunctivity-disjunctivity, succorance-autonomy, impulsion-deliberation, etc., so that the total number of needs to which scores can be attached is 42 rather than 30. A preliminary manual for the Activities Index describes the test in detail (**8**).

A corresponding test for describing college environments, called the College Characteristics Index, was subsequently constructed. It consists of 300 statements about college environments to which responses of "True-False" are given. The statements are organized into 30 ten-item scales, with a press scale for each need scale that was included in the Activities Index. The following kind of questions guided the writing of items: what might be characteristic of an environment which exerted a press toward order, or toward autonomy, or toward nurturance, or understanding, or play, etc? Stated in another way, what might there be in a college environment which would be satisfying to or tend to reinforce or reward an individual who had a high need for order, or autonomy, or nurturance, or understanding, or play, etc.? The items themselves are statements about college life. They refer to the curriculum, to college teaching and classroom activities, to rules and regulations and policies, to student organizations and activities and interests, to features of the campus, etc.

Sample items from corresponding Need and Press scales will illustrate the parallelism.

A need for Order would be inferred from liking such activities as: "Arranging my clothes neatly before going to bed. Hav-

[1] Portions of this research have been financed by grants from the College Entrance Examination Board and the Carnegie Corporation.

ing a special place for everything and seeing that each thing is in its place. Keeping a calendar or notebook of the things I have done or plan to do." What might such a person like to find in a college environment or what features of a college environment might be rewarding or frustrating to such a need? The following items from the press scale for Order might be relevant: "Faculty members and administration have definite and clearly posted office hours. In many classes students have an assigned seat. Professors usually take attendance in class."

On the need scale for Impulsion-Deliberation, a high score for Impulsion results from liking such activities as: "Being in a situation that requires quick decisions and actions. Doing things on the spur of the moment. Doing whatever I'm in the mood to do." Thus, a college environment which has a press toward impulsiveness might be a place where: "Most students don't decide what courses to take until the time of registration. Students often start projects without trying to decide in advance how they will develop or where they may end. Spontaneous student rallies and demonstrations occur frequently."

A high need for Energy is inferred from liking such activities as: "Taking up a very active sport. Having something to do every minute of the day. Giving all of my energy to whatever I happen to be doing." The needs of such a person might be expected to find fulfillment and satisfaction in a college environment where: "There is an extensive program of intramural sports and informal athletic activities. Student gathering places are typically active and noisy. Class discussions are typically vigorous and intense."

Just as needs are inferred from the characteristic modes of response of an individual, so press are reflected in the characteristic pressures, stresses, rewards, conformity-demanding influences of the college culture. Operationally, press are the characteristic demands or features as perceived by those who live in the particular environment. To each statement in the College Characteristics Index the person who takes the test answers true if he believes it is generally characteristic of the college, is something which occurs or might occur, is the way people tend to feel or act; and he answers false if he believes it is not characterisic of the college, is something which is not likely to occur, is not the way people typically feel or act.

A Study of Five Institutions

A first draft of the College Characteristics Index was administered in May, 1957, to groups of students at five institutions and to smaller groups of faculty members at four of the five institutions. In all, 423 students and 71 faculty members responded to the instrument. Neither student nor faculty groups were representative samples. Most of the students were upperclassmen and most of the faculty members were in the upper academic ranks. It was argued that if a dominant press really exists in a particular environment almost any group of people living in the environment would probably identify it. The testing program was, in any case, intended only as a preliminary try-out of the model from which some information would be gained about the types of items, the possible reliability and validity of the scales, and the potential utility of this approach to measuring college characteristics.

The five institutions, although not identified here, were selected because observers would probably agree that they are rather different from one another, with the selection of colleges thus providing some evidence about the construct validity of the test. One was a large Midwestern state university. The second was a large Midwestern private university. The third institution was a large Eastern private uni-

versity. The fourth was a moderate-sized Eastern private college for men. The fifth institution was a publicly supported college in the metropolitan New York area.

Test Results

Saying that a particular press is or is not characteristic of an institution is an arbitrary matter. There exist no conventions or experience to guide the decision. The basis for the tentative decisions that were made, together with other statistical information about the scales and the items are given in the following paragraphs.

In examining means, or a profile on which mean scores are plotted, one naturally looks for what appear to be high and low points. And in examining variances of distributions, in the context of press identification, one naturally looks for some concentration of scores or a low variance, suggesting a consistency of impression, rather than a wide dispersal of scores, suggesting divergent impressions about the press.

Table 1 shows the means and standard deviations on each of the press scales, computed from the students' responses at the five institutions studied. Each of these 30 scales has a maximum possible range of 0 to 10. The median of these mean scores is approximately 5.5. The median of the standard deviations is approximately 1.7. Putting this information together, one might suggest that, for the five institutions represented, a fairly reasonable definition of a noticeable press (or its absence) would be one which required a mean score falling in the upper or lower one fourth of the total distribution. Mean scores of 6.6 or higher and mean scores of 4.4 or lower would thus be suggestive of a press. In Table 1, means which meet this criterion are in italics.

A corresponding table showing faculty means and sigmas is not presented, mainly because the number of cases is quite small—*N*s being 25, 20, 11, and 15. Also, from an educational view, it may be argued that the effective press of an environment is what the students say it is, not what the faculty say, or what the catalogue says, etc. Nevertheless, in the four colleges where both student and faculty responses to the CCI were obtained it is of some pertinence to note their similarity. Table 2 shows a distribution of the differences between the means of students' and faculty responses in each institution and also the differences between faculty and student responses to individual items. Here it is evident that most of the differences are fairly small. In nearly half of all the comparisons between faculty and student mean scores the difference between the two was less than half a point. Three fourths of all the differences were less than 1.00. Differences between faculty and student responses to individual items are grouped in three arbitrary categories. Thus, 44% of all the items answered by students and faculty were answered within a range of agreement of 10 percentage points or less; and on 12% of the items the percentage for the students' answer differed by 30 points or more from the percentage for the faculty's answer. Considering the number of cases in all groups, one can be confident that differences of 10 points or less are in all instances merely chance differences, and that differences of 30 points or greater may in most instances be significant differences at least at the 5% level of confidence. It appears then, that some of the differences in the middle category of the table are chance differences and some are significant differences. One might estimate that, over all, about three fourths of the items were answered in tolerably good agreement by both students and faculty, and that perhaps as many as one fourth of the item responses represent divergent views between students and faculty in characterizing the institution.

TABLE 1

MEANS AND STANDARD DEVIATIONS ON THE VARIOUS PRESS SCALES FROM STUDENTS' RESPONSES AT FIVE INSTITUTIONS

Press Scales	College A (N = 100)		College B (N = 44)		College C (N = 100)		College D (N = 68)		College E (N = 111)	
	$\bar{M}$	SD	$\bar{M}$	SD	$\bar{M}$	SD	$\bar{M}$	SD	$\bar{M}$	SD
Abasement	4.0	2.2	3.6	1.7	3.3	1.9	3.4	1.7	4.3	2.1
Achievement	4.9	1.9	5.3	1.4	4.4	1.9	5.4	1.9	5.5	1.8
Adaptiveness	6.3	1.8	7.9	1.4	6.9	1.5	6.3	1.5	5.3	1.6
Affiliation	5.4	2.0	5.6	2.2	5.8	1.7	4.2	2.1	3.8	2.0
Agression-Blameavoidance	4.4	1.9	5.8	1.4	5.6	1.6	4.7	1.6	3.5	2.0
Change-Sameness	5.6	1.3	7.1	1.4	5.1	1.7	5.1	1.4	4.8	1.4
Conjunctivity-Disjunctivity	6.7	2.0	5.8	2.3	7.4	1.7	5.8	2.1	6.9	1.9
Counteraction	6.1	1.4	6.5	1.4	5.0	1.5	5.4	1.6	5.9	1.4
Deference	5.6	1.7	5.1	1.4	5.6	1.7	5.4	1.4	5.5	1.5
Dominance	5.5	1.5	3.5	1.5	4.9	1.7	5.7	1.6	4.3	1.8
Ego-Achievement[a]	6.9	2.0	5.2	1.5	5.6	1.7	6.9	1.8	4.9	2.2
Emotionality-Placidity	6.3	1.5	5.8	1.6	6.3	1.5	5.9	1.9	5.3	1.8
Energy-Passivity	5.1	1.9	6.0	1.7	4.4	1.8	5.0	1.6	4.8	1.7
Exhibitionism	6.7	1.7	5.8	1.6	6.2	1.4	6.3	1.5	5.6	1.8
Fantasied Achievement[b]	6.0	1.8	6.5	1.6	5.7	1.6	4.6	1.7	5.7	1.6
Harmavoidance	4.7	1.5	4.8	.9	3.4	1.3	3.9	1.3	5.5	1.5
Humanism[c]	5.7	1.9	8.5	1.3	6.0	1.8	6.8	1.8	6.6	2.0
Impulsion-Deliberation	3.5	1.5	4.3	1.5	4.1	1.5	4.2	1.3	3.3	1.4
Narcissism	5.8	1.5	3.2	1.3	3.8	1.6	5.3	1.7	5.5	1.7
Nurturance	7.4	1.6	5.8	1.8	6.7	1.7	6.9	1.7	6.5	1.7
Objectivity-Projectivity	6.6	1.9	8.1	1.4	7.4	1.9	6.7	1.9	6.4	2.1
Order	7.3	1.5	3.3	1.5	5.8	1.4	6.2	1.4	7.1	1.3
Play	7.4	1.5	3.8	1.6	6.1	1.4	7.8	1.6	5.9	1.9
Pragmatism[d]	4.6	1.6	1.7	1.1	4.9	1.6	4.7	1.5	3.8	1.6
Reflectiveness[e]	5.5	1.9	8.0	1.4	4.9	1.9	5.5	1.9	5.6	1.9
Scientism[f]	6.3	1.9	8.4	1.5	5.1	2.0	6.5	2.0	6.5	1.7
Sentience	5.3	1.8	6.5	1.5	3.5	1.6	5.2	2.0	5.0	1.8
Sex	6.2	1.4	4.9	1.5	5.7	1.3	7.0	1.3	4.7	1.3
Succorance-Autonomy	5.8	1.3	4.6	1.6	5.2	1.4	5.1	1.9	5.1	1.6
Understanding	4.8	1.8	8.6	1.2	5.1	1.7	4.7	1.6	5.8	1.8

[a] Derived from Exocathection-Intraception.
[b] Derived from Ego Ideal.
[c] Derived from Endocathection-Extraception: Social Sciences and Humanities.
[d] Derived from Exocathection-Extraception.
[e] Derived from Endocathection-Intraception.
[f] Derived from Endocathection-Extraception: Natural Sciences.

Reliability

Test-retest data are not available in the present study. Conventional estimates of reliability from a single administration may not be appropriate for an instrument on which one hopes to find skewed distributions and minimal dispersion of scores. Faculty-student agreement within the same institution is of some relevance but one might argue that the perceptions of these two groups could differ on individual items or scales, yet each could be reliable within itself.

Item analysis data are more directly relevant to the reliability of scales. An item analysis was made of each scale, separately for each of the five institutions, using the students' responses. Ebel's (1)

simplified method was used. Each of the 30 scales was thus subjected to item analysis in five different samples, and since there are 10 items in each scale, the total number of item discrimination indexes obtained was 1500. Of these 1500 discrimination indexes, 1% were negative, 18% fell between .0 and .19, 30% fell between .20 and .39, and 51% were .40 or higher. In other words, 81% of the items had, on the average, moderate to high discrimination in their respective scales.

Perhaps the most important approach is one which treats reliability and validity as inseparable and deals with the instrument as a whole. For example, do different people characterize the institution in the same way? This involves the reliability of profiles, with all their interrelationships. As a first approximation of this, the rank order of mean scores from the students' responses can be compared with the rank order of mean scores from faculty responses within the same institution. Thus, do these groups see the institution in relatively the same pattern? For the two colleges (Colleges B and C) which had the largest number of faculty respondents, these rank order correlations were .96 and .88. Correlations were not computed for the other two colleges where faculty mean scores would be based on *N*s of 15 and 11.

TABLE 2

SUMMARY OF DIFFERENCES BETWEEN FACULTY AND STUDENT RESPONSES WITHIN EACH OF FOUR INSTITUTIONS

Mean Differences in Scale Scores	Number of Differences	Cumulative Per Cent
.0 to ±.49	57	48%
±.50 to ±.99	34	76%
±1.00 to ±1.49	23	95%
±1.50 to ±1.99	6	100%
	120	
Percentage Differences in Responses to Individual Items		
10 percentage points or less	528	44%
Between 11 and 29 points	528	88%
30 points or more	144	100%
	1200	

Validity

To illustrate the sort of interpretation and description of a college environment which can be derived from the CCI, the press of two colleges are presented. The comments are based entirely on the arbitrary definitions of what levels of scores constitute a press (which was explained earlier); and the nature of these press is further illustrated by citing some of the specific items which most clearly define them. No estimate is presently available of the validity of these descriptions against a systematic outside criterion. But the descriptions show quite clearly that these are very different environments, and that the test is therefore capable of revealing some sharp distinctions between colleges which qualified observers would expect to be different. The evidence is therefore relevant to the property of validity. Detailed data, and descriptions of the other three colleges, may be found elsewhere (7).

College A.

The major press in College A are toward orderliness and friendly helpfulness, with overtones of spirited social activity. This is suggested by high scores on the scales for Order, Objectivity, Conjunctivity, Nurturance, Play, Ego Achievement, Exhibitionism, and by low scores on the scales for Abasement, Impulsion, and Aggression.

The stress on Order, Deliberation (opposite of Impulsion), and Conjunctivity is indicated by such highly shared observations as the following: students have assigned seats in some classes, professors

often take attendance, papers and reports must be neat, buildings are clearly marked, students plan their programs with an adviser and select their courses before registration, courses proceed systematically, it is easy to take clear notes, student activities are organized and planned ahead.

Within this orderliness, student life is spirited and a center of interest. For example, big college events draw lots of enthusiasm, parties are colorful and lively, there is lots to do besides going to classes and studying, students spend a lot of time in snack bars and in one another's rooms, and when students run a project everyone knows about it.

At the same time, amid this student-centered culture, there is a stress on idealism and service. Students are expected to develop an awareness of their role in social and political life, be effective citizens, understand the problems of less privileged people, be interested in charities, etc.

The total picture of the environment, then, is one of high social activity, esprit de corps, and enthusiasm combined with an emphasis on helping others and idealistic social action and all within a fairly well understood set of rules and expectations which are deliberative and orderly. One would expect some of the explicit objectives of such an institution to stress personal and social development, idealism and social action, and civic responsibility.

College B.

Here the dominant press of the environment fall in the theoretical-intellectual category—Reflectiveness, Humanism, Scientism, Understanding, and Objectivity. This dominant press occurs in an environment also characterized by Change, nondefensive acceptance of criticism (Adaptiveness), and by resistance to any abject acceptance of criticism or presumed low status (Abasement). Moreover, on two of the scales which defined a high press at College A (Play and Order), the press at B is exactly opposite. There is further, a minimum of importance to social status or manipulation for tangible ends (Pragmatism), preoccupation with self and personal appearance (Narcissism), and bossing or directing others (Dominance). There is, however, a generally consistent and high press toward Deliberation or planning and thinking ahead.

It is clear that the most pervasive press is directed toward the pursuit of understanding for its own sake, abstract and unencumbered by requirements for practical utility or social action.

The theoretical-intellectual press of the environment at College B is more specifically suggested by the following observations with which, generally, more than nine tenths of both faculty and students agree: there are excellent library resources in natural science and social science, a lecture by an outstanding philosopher or scientist would draw a capacity audience, many students are planning graduate work or careers in science or social science, there are many opportunities for students to see and hear and criticize modern art and music, reasoning and logic are valued highly in student reports and discussions, students who spend a lot of time in a science laboratory or in trying to analyze or classify art or music or in seeking to develop a personal system of values are not regarded as odd, scholarship and intellectual skills are regarded as more important than social poise and adjustment, there is time for private thought and reflection, one need not be afraid of expressing extreme views, the faculty and administration are tolerant and understanding in interpreting regulations.

In contrast with College A, students at B do not have an assigned seat in class, professors do not take attendance, students are likely to study over the weekend, big college events draw no great enthusi-

asm, and the place is not described as one where "everyone has a lot of fun."

Moreover, student leaders have no special privileges, family status is not important, students are not much concerned about personal appearance and grooming, and an intellectual is not an "egghead."

And finally, exams are not based on factual material from a textbook, classes are not characterized by recitation and drills, grade lists are not publicly posted, students are not publicly reprimanded for mistakes, student organizations are not closely supervised, students tend to stay up late at night, work all the harder if they have received a low grade, and if confronted with a regulation they do not like they will try to get it changed.

One would expect the explicit objectives of such an institution to stress the acquisition of knowledge and theory, critical judgment and independence, and a sense of the significance of intellectual life.

Other Differences Between Colleges

That College A and College B are quite different environments can be suggested statistically as well as in the descriptive profiles just presented. The rank order of mean scores for College A correlates, for example, .06 with the rank order of mean scores for College B. One can also estimate, without computing, the significance of the differences between the mean scores of College A and College B on each of the 30 scales. Given *Ns* of 100 and 44, and standard deviations as large as 2.00, a difference of 1.00 is significant well beyond the 1% level of confidence. On at least 22 of the 30 scales, the differences between means for the two schools are significant. Considering all five schools, it is clear that on every scale except one (Deference) there are significant differences between two or more of the mean scores.

Another indication of differences can be obtained from noting the differences between the percentages of students at Colleges A and B answering each item according to the key. In the middle range of percentages a difference of 20 points, with *Ns* of 100 and 44, is always significant beyond the 1% level. On this conservative basis there were 172 of the 300 items on which the percentages for College A differed significantly from College B.

The point to these observations is merely to suggest that the first trial run of the CCI produced many results which clearly differentiated among the environments or press of the five colleges. The actual items in the CCI are not reproduced here because there seems little virtue in printing a first draft. The content of many items has been, of course, indicated in the descriptive profiles of the two colleges and in the earlier paragraphs which noted the parallel structure of Need items and Press items.

Conclusions and Implications

After completing the preliminary studies reported thus far, a revised form of the College Characteristics Index was produced in which 58% of the original items were retained, 13% were slightly modified, and 29% of the items were new. The revised form, at the time of writing this article, has been administered to approximately 1200 students distributed among more than 30 institutions. Further research and more intensive analyses have been planned. Before commenting on specific plans, however, certain broad values and implications in this psychological approach to the measurement of college environments are suggested.

One potential value, for example, is in institutional self-analysis. Administrators and faculty members should be able to learn something useful about the dynamics of the college environment from studying students' responses to the College Characteristics Index. Institutional press should have some clear relationship to institu-

tional purpose. The objectives of a college are formal or explicit statements of intent: they indicate the directions in which a college means to influence the behavior of students. They find expression in curricula, practices, services, policies, and other aspects of the college environment. The press, as measured by the method described, constitute what Stern, Stein and Bloom (**9**) have referred to as an operational definition of objectives, or the implicit influence of the environment upon the students. Implicit press and explicit objectives should reinforce one another, for an institution should operate in reality the way it means to operate in theory. Consequently, a serious lack of congruence between implicit press and explicit objectives would suggest to faculty members and administrators that certain aspects of the environment ought to be changed in order to make the total impact of the institution more consistent or more effective. Pace (**6**) has commented elsewhere on the disintegrative effect of discrepancy between stated objectives and actual practices.

Some aspects of an environment can be changed more readily than others. The College Characteristics Index provides some direct indications of the psychological implications of various policies and practices. Roughly, one fourth to one third of the items in the Index state specific practices which an administration or faculty could more or less easily change if they did not like the implications. For example, being able to drop a course in which one is having difficulty, or to substitute another course for one which has been failed, is associated with Counteraction; insisting that students' reports or papers be neat, or giving students an assigned seat in class, or taking attendance regularly, is associated with Order. As the relationships among press variables and between these variables and institutional objectives as well as personal needs are established, the significance of such specific practices will become clarified. Other items in the Index are more indirect in their implications about the effect of various policies or practices. But the clues can be investigated and can thus be the starting point for serious discussions about the impact of the environment on the student and the relation of this impact to the intended objectives.

Another set of implications from this approach to describing college environments relates to the problem of assessment and prediction. Assessment studies have often failed because the situations or environments in which assessed modes of behavior were supposed to occur have been inadequately described or differentiated. The interaction between person and environment was not successfully predicted because the environment was not measured as analytically and systematically as the person. The whole field of college prediction studies provides a good example. The criterion is typically a grade point average. No fundamental improvement in predicting against this criterion has been made in the last 25 years. Prediction studies should be concerned with performance in the environment as a whole. The complexity of relationship between person and environment is inevitably obscured by the simplified and often inappropriate symbolism of correlation between scholastic aptitude test and grade point average. The press of a college environment represents what must be faced and dealt with by the student. It is possible that the total pattern of congruence between personal needs and environmental press will be more predictive of achievement, growth, and change than any single aspect of either the person or the environment.

It will be a long time before admissions officers or guidance counselors can benefit from the results of these more complex analyses. This requires the establishment

of known relationships between kinds of persons and kinds of environments. But conceivably, advisers within an institution may be better able to help students find an effective and rewarding role within the operative environment of the college, or to see more clearly the ways in which environments need to be modified if different kinds of students are to grow within them most effectively.

As further steps toward refining and exploring the potential of this needs-press concept, as exemplified by the Activities Index and the College Characteristics Index, the following studies, among others, are in progress:

1. Statistical studies of the instrument (CCI), including test-retest reliabilities, correlation matrix of all scale scores, factor analysis, item covariances.

2. Comparative studies of types of items, subjective or impressionistic versus relatively objective or factual.

3. Studies of the relations between students' needs scores and the corresponding perception of press in the environment.

4. Analysis of perception of press among various sub-cultures within a complex institution.

5. Significance of congruence between Need and Press in determining successful performance and/or satisfaction in the college environment.

Summary

The present article has suggested that a college environment may be viewed as a system of pressures, practices, and policies intended to influence the development of students toward the attainment of important goals of higher education. The first draft of an instrument for measuring these influences systematically has been constructed. Data analyzed thus far reveal significant differences in the press of different college environments. The instrument itself, which has subsequently been revised, appears to be promisingly reliable and valid. In the long run, research which this type of instrument makes possible should increase understanding of the ways in which institutions make their impact upon students and provide a broader conceptualization for evaluating the effectiveness of higher education.

REFERENCES

1. Ebel, R. L. Procedures for the analysis of classroom tests. *Educ. psychol. Measmt.*, 1954, **14**, 277–282.
2. Edwards, A. L. *Edwards Personal Preference Schedule*. New York: The Psychological Corp., 1954.
3. Gardner, E. F., & Thompson, G. G. *Social relations and morale in small groups*. New York: Appleton-Century-Crofts, 1956.
4. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
5. Murray, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
6. Pace, C. R. Educational objectives. The integration of educational experiences. *Yearb. nat. Soc. Stud. Educ.*, 1958, **57**, Part III. Pp. 69–83.
7. Pace, C. R., & Stern, G. G. *A criterion study of college environment*. Syracuse: Syracuse Univer. Res. Inst., Psychol. Res. Center, 1958.
8. Stern, G. G. *Preliminary manual: Activities Index; College Characteristics Index*. Syracuse: Syracuse Univer. Res. Inst., Psychological Res. Center, 1958.
9. Stern, G. G., Stein, M. I., & Bloom, B. S. *Methods in personality assessment*. Glencoe, Ill.: Free Press, 1956.

Received June 5, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 49, No. 5, 1958

PRESCHOOL IQs AFTER TWENTY-FIVE YEARS[1]

KATHERINE P. BRADWAY
Stanford University

CLARE W. THOMPSON[2]
State College of Washington

AND RICHARD B. CRAVENS[2]
Department of Health, Territory of Hawaii

In 1931, 212 children in the San Francisco Bay Area between the ages of 2 and 5½ years were given tests which later constituted Forms L and M of the 1937 Revision of the Stanford-Binet. Careful methods were developed and adhered to in the selection of these children to avoid biased sampling and to insure a representative group, because they were the California sample of the nationwide standardization of this revision (**9**, pp. 12–15, 18; **7**, pp. 6–7, 36–37).

Ten years later, in 1941, 138 of these children still in the area were administered Form L of the Stanford-Binet.[3] Thirteen of the original group were not included because as two-year-olds they had missed so many items at the lowest test level that the obtained mental age was a maximum expression of their intelligence, the true estimate not being determinable. Another 61 could not be located.

In 1956, 111 of the 1941 group were located and given Form L of the Stanford-Binet and the Wechsler Adult Intelligence Scale. Nine *Ss*, although located, refused to participate; one *S* was in a mental hospital and not accessible for testing; another *S* was located in a community where no provisions could be made for testing; the remaining *Ss* could not be located. A detailed report of the results of this follow-up and their theoretical implications is in the process of preparation. This is being interrupted now to present a summary of the actual IQ changes over the 25-year period because of the pending appearance of a new revision of the Stanford-Binet[4] and the appropriateness of having these retest data on the standardization group available at the same time.

SUBJECTS AND PROCEDURE

The *Ss* in the present study were administered Forms L and M of the Revised Stanford-Binet Scale 25 years previously, in 1931, in connection with its standardization. They were also examined with Form L of this same scale in 1941 (**3**, **4**, **5**). No interim contacts had been made. A comparison of the 1941 retest group of 138 *Ss* with the total standardization group at these ages showed some upward selection, as the initial mean composite IQ of the retest group was 109.2 compared with 105.4 for the standardization group. Most of this difference was concentrated in the younger age groups,

[1] This investigation was supported by Research Grant M-1273 from the National Institutes of Health, Public Health Service.

[2] At Stanford University when participating in this research.

[3] The authors wish to express their debt of gratitude to Maud A. Merrill and the late Lewis M. Terman for having made their standardization data available for follow-up. They also wish to express their appreciation for the continued interest of Quinn McNemar, who has served as consultant for both follow-ups.

[4] Personal communication from Maud A. Merrill.

where the children for whom the test was too difficult had been eliminated. The distributions of paternal occupational level for the retest and standardization groups were similar at all ages.

The loss of *Ss* between the 1941 and 1956 testings resulted in further selection with respect to initial mean IQ, which rose to 112.8 for the 1956 group with the elimination of 27 *Ss*, 25 of whom were not located or refused to participate. Apparently availability for follow-ups was positively related to intelligence, although again the greatest discrepancy was for the group which had been two years of age at the initial testing.

In the current testing of 1956, Form L of the Revised Stanford-Binet and the Wechsler Adult Intelligence Scale were administered. (In one case the WAIS was incomplete. In another the 1931 IQ is not valid; this case was included only in comparisons with the 1941 results.) Of the 111 *Ss*, 98 were tested by Cravens; two were tested by Bradway, who had administered all tests in the 1941 follow-up; the remaining 11 *Ss* were tested by psychologists at colleges and universities located near the *Ss*' current place of residence.[5] Mobility was less than had been anticipated: of the 122 *Ss* located, only 24 lived more than 40 miles from where they had lived in 1931.

The retest group was comprised of 52 men and 59 women. This is the same proportion as in the 1941 testing. Inasmuch as no sex differences in test results were found in the standardization (**9,** p. 34), the results in the present study are given for the total group, combining both sexes.

[5] The authors wish to acknowledge their indebtedness to the following psychologists who participated in this way: Steven K. Abe, Walter H. Brackin, Maurice Deigh, Robert Gray, George S. Ingebo, W. B. Lemmon, Jeanne Reitzell, Gordon and Virginia Riley, Alwyn Sessions, Paul S. Spitzer, and Helene R. Veltfort.

TABLE 1

OBTAINED STANFORD-BINET IQs BY YEAR OF TESTING

	Mean IQ	SD
1931[a]	112.8	15.9
1941[b]	112.3	16.4
1956[b]	123.6	15.0

[a] Composite Form L and Form M.
[b] Form L.

RESULTS

The degree of relationship between the initial composite IQ (Forms L and M) of these *Ss* when they were in the age range from 2.0 to 5.5 (mean age 4.0) and Form L IQ when they were in the age range from 26.5 to 32.2 (mean age 29.5) is expressed by a Pearsonian *r* of .59. This compares favorably with the *r* of .65 found for the same group in the first follow-up after only 10 years. The correlation between the 1941 and the 1956 testings is .85.

All but 19 of the 110 *Ss* showed a higher IQ on the 1956 testing than they did in 1931. This was not limited to any one segment of the IQ range. The mean increase was 10.8 points. The results in Table 1 show that this increase occurred between 1941 and 1956, that is between the early adolescent years and adulthood, since the mean IQs in 1931 and 1941 were approximately the same. The mean IQ increase from 1941 to 1956 was 11.3 points. The increases from 1931 to 1956 and from 1941 to 1956 are, of course, highly significant statistically with CRs of 8.1 and 13.7 respectively. The 1956 WAIS mean IQ of 108.9 ± 11.0 more nearly approximates the 1931 and 1941 Stanford-Binet mean IQs than does the 1956 Stanford-Binet.[6]

[6] The correlational values for the 1956 WAIS for this group are similar to those for the 1956 Stanford-Binet. Pearsonian *r* with 1931 Stanford-Binet is .61; with 1941 it is .80. The correlation between the two 1956 tests is .83.

These data are interpreted as invalidating the assumption that intelligence stops increasing at 16 years, on which was based the standard computation of an adult Stanford-Binet IQ. This is consistent with the findings of other recent investigators. Bayley (**1**) in the Berkeley Growth Study found increases in Stanford-Binet scores of the same *Ss* up to the age of 17 years (the most recent administration), and on the Wechsler-Bellevue up to the age of 21 years (most recent administration). Moreover, a few 25-year scores available at the time of writing indicated that the ceiling of mental growth had not yet been reached by these *Ss*. The interpretation of these results, of course, must take account of the possibility of practice effects from frequent retesting. Owens (**8**) found an increase on Army Alpha scores by *Ss* who took the test initially at 19 years and again at 50 years. Bayley and Oden (**2**) in the most recent examination of Terman's Gifted Children found increases in scores on the Concept Mastery Test for *Ss* covering a total range from age 20 to age 50 years, retested after a 12-year interval.

It seems obvious that a correctional factor of some sort should be applied to the adult Stanford-Binet IQs. Decision on the nature of such a correction for purposes of analyzing data of the present study is being deferred until the pending publication of the new revision. It may be noted, however, that whereas the IQs of 42% of the present group changed more than 10 points from 1931 to 1941, the IQs of 60% changed more than 10 points from 1931 to 1956. If the IQ is to be meaningful as an index, the average gain should be zero as it was between 1931 and 1941. When the IQs are equalized by taking account of the mean increase in IQ and subtracting 11 points from the 1956 value, the frequency of changes greater than 10 points is reduced to 42%. Similarly, 22% of the present group changed more than 15 points from 1931 to 1941, whereas 41% changed more than 15 points from 1931 to 1956. Equalizing the IQs as described above, the latter is reduced to 22%. From 1941 to 1956 the equalizing of the values reduces those changing more than 10 points from 58% to 18% and those changing more than 15 points from 28% to 7%.

Discussion

Beyond the problem of determining the optimal method to use in computing adult indices of intelligence is the question of what implications these data have for the nature of the mental growth curve and the theoretical question of terminal age of intellectual growth. A choice of method to use in arriving at the most useful index of adult intelligence requires the consideration of many factors. Wechsler's use of IQs computed by setting a mean of 100 and a standard deviation of 15 for each age is one solution. The central problem, however, is that intelligence, like so much of human behavior, is multifaceted. The point of view from which the question is approached depends upon what aspect is to be explored. Are we interested in an index showing one's place relative to others of the same age group, an index of one's location on a scale of development and decline, an assessment of one's various kinds of mental processes? Inasmuch as different abilities reach maturity at different ages and begin to decline at different ages, the particular abilities examined become deciding factors. For example, Corsini and Fassett (**6**) found from the testing of 1072 adults on the Wechsler-Bellevue that general intelligence declines from early to late maturity only if visual and motor factors are tested, and increases if the items are dependent on continued learning. We have in progress further research directed at determining what factors are associated with fluctuation in measured intelligence throughout the life span and the course of development and decline of different kinds of intelligence.

SUMMARY

Of the 212 preschool children living in the San Francisco Bay Area who composed part of the standardization group of the 1937 Revision of the Stanford-Binet, 111 were retested as adults 25 years later. The sample was somewhat biased, with a tendency for *Ss* at the lower IQ levels to be more difficult to locate or to refuse to participate. Thus the initial mean IQ of those retested was 112.8 as compared with 105.4 for the total group.

The Pearsonian *r* between adult and preschool Stanford-Binet IQs over the 25-year period is .59. The mean IQ showed a rise of 10.8 points in 25 years. This rise occurred in the years after early adolescence, there being no increase in mean IQ between preschool and adolescent testings. This is interpreted as invalidating the assumption that intelligence stops increasing at 16 years. Further research is in progress.

REFERENCES

1. BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, **10**, 805–818
2. BAYLEY, NANCY, & ODEN, MELITA H. The maintenance of intellectual ability in gifted adults. *J. Geront.*, 1955, **10**, 91–107
3. BRADWAY, KATHERINE P. IQ constancy on the Revised Stanford-Binet from the preschool to the junior high school level. *J. genet. Psychol.*, 1944, **65**, 197–217
4. BRADWAY, KATHERINE P. An experimental study of factors associated with Stanford-Binet IQ changes from the preschool to the junior high school. *J. genet. Psychol.*, 1945, **66**, 107–128
5. BRADWAY, KATHERINE P. Predictive value of Stanford-Binet preschool items. *J. educ. Psychol.*, 1945, **36**, 1–16
6. CORSINI, R. J. & FASSETT, K. K. Intelligence and aging. *J. genet. Psychol.*, 1953, **83**, 249–264
7. MCNEMAR, Q. *The Revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin, 1942
8. OWENS, W. A. Age and mental abilities. *Genet Psychol. Monogr.*, 1953, **48**, 3–54
9. TERMAN, L. M., & MERRILL, MAUD A. *Measuring Intelligence*. Boston: Houghton Mifflin, 1937.

Received June 18, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol., 49, No. 5, 1958

THE ADEQUACY OF "MEANING" AS AN EXPLANATION FOR THE SUPERIORITY OF LEARNING BY INDEPENDENT DISCOVERY[1]

BERT Y. KERSH

University of Oregon

The hypothesis that learning through independent discovery is superior to learning by rote is well supported by existing research evidence (**5, 6, 7**). More recently, however, evidence has been published which suggests that attempts to direct the learner in the discovery process may also be successful without loss in retention or transfer (**2, 3**).

One reasonable explanation for the superiority of both the independent discovery and the directed discovery processes is that learning under either condition is more meaningful than in the case where the learner simply memorizes answers. Meaning is used in the following report in the cognitive sense of understanding or organization. More precisely, through the discovery process, in which the learner is forced to rely on his own cognitive capacities, he becomes cognizant of the relationships of the learning task to his previous experience, or to the pattern of relationships among the elements of the task. The superiority of such meaningful learning over rote learning is also well supported by research evidence (**1, 4**).

If meaningful learning is the key concept, it should make no difference whether learning occurs with or without direction, so long as the learner becomes cognizant of the essential relationships. However, it is very likely that some procedures of learning may be superior to others simply because they are more likely to cause the learner to become cognizant of the relationships.

The purpose of this research was to study the process of learning tasks involving arithmetical and geometrical relationships in order to determine whether or not the superiority of the discovery and directed-discovery procedure is adequately explained in terms of "meaningful learning," and, if not, to discover a more adequate explanation.

DESCRIPTION OF THE TASKS

Each *S* had the task of learning the following two rules of addition.

1. *The Odd-numbers rule.* The sum of any series of consecutive odd numbers beginning with one is equal to the square of the number of figures in the series. (For example, 1, 3, 5, 7, is such a series; there are four numbers, so 4 times 4 is 16, the sum.)

2. *The Constant-difference rule.* The sum of any series of numbers in which the difference between the numbers is constant is equal to one-half the product of the number of figures and the sum of the first and last numbers. (For example, 2, 3, 4, 5, is such a series; 2 and 5 are 7; there are four figures, so 4 times 7 is 28; half of 28 is 14 which is the sum.)

These rules can be learned by simply memorizing the task procedure as above. On the other hand, the learner can become cognizant of certain relationships to geometrical and arithmetical concepts which the two rules involve, in which case his learning will be more meaningful.

Consider first the Odd-numbers rule. In the first place, it is possible to relate the arithmetical concept of "squaring a number" to the geometrical concept of

[1] This study was supported by research grants from The Graduate School of the University of Oregon, Eugene, Oregon. Research assistance was ably provided by Harold K. Kozuma and Arthur O. Motz. An abbreviated version of this report was read at the joint meeting of the AERA and CERA in San Francisco, March 1958, under a different title.

"square." For example, "4^2" may be conceptualized as the arithmetical counterpart to the geometrical concept of a square having 4 units to the side.

Now, it can readily be shown that any series of odd numbers beginning with one can be rearranged in the form of a geometric square, as is illustrated in Fig. 1. In Fig. 1, each row of X's represents a number—the first row is "1," the second row is "3," and so on. The X's which are left outside the square can be fitted into the spaces indicated by the dashes inside the box.

The rule can also be related to the concept of the arithmetical average. The mean of any such series is the middle number; and the mean is always equal to the number of figures in the series. So the *mean times the number* is equivalent to the *number squared*. This relationship is suggested in Fig. 2.

Consider now the Constant-difference rule. The arithmetical and geometrical relationships involved in this rule are suggested in Fig. 3 and 4. Each figure essentially represents the original series inverted and added across; and it becomes apparent that the sum of the first and the last number, the sum of the second and the second-from-the-last number, and so on, is the same. The sum of the original series is, of course, one half the sum of the column of sevens formed in Fig. 3, and half the area of the rectangle thus formed in Fig. 4.

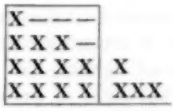

FIG. 1.

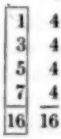

FIG. 2.

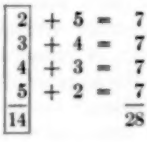

FIG. 3.

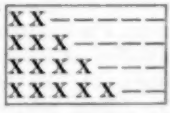

FIG. 4.

HYPOTHESES

The experiment was designed to test the following hypotheses:

1. It makes no difference in terms of retention or transfer effects whether the learner discovers the relationships which are essential to the understanding of a cognitive task independently or with external direction.

2. It is more probable that a learner becomes cognizant of the relationships which are essential to understanding a cognitive task when his attention is directed to the relationships than when his attention is directed to the task procedure alone, or when he is required to learn independently.

3. It follows from the second hypothesis that it is more probable that what is learned is remembered longer and transferred more effectively when the learner's attention is directed to the essential relationships than when his attention is directed to the task procedure alone, or when he is required to learn independently.

4. In the learning of tasks which involve both arithmetical and geometrical relationships, the learner most probably becomes cognizant of the former when conventional Hindu-Arabic symbols are used in directing his learning, and of the latter when more nearly iconic symbols are used.

PROCEDURE

One group of *Ss*, called the "no-help"

group, was required to discover the rules without help. A second group, called the "direct-reference" group, was given some direction in the form of perceptual aids, such as those illustrated in Fig. 1–4 above, with accompanying verbal instructions which directed their attention to the perceptual aids. A third group, called the "rule-given" group, was told the rules directly and was given practice in applying them, without any reference to the arithmetical or geometrical relationships. The three procedures were called, as a group, the "teaching treatments."

In addition, there were two treatments called the "number treatments," which consisted simply of presenting the problems in either the more nearly iconic form, called the "X-form," or the conventional Hindu-Arabic form, called the "A-form."

The three teaching treatments and the two number treatments were combined to form six treatment combinations in a 2 × 3 factorial design. The treatment combinations were identified as follows: No help, A-form (No-help A); No help, X-form (No-help X); Direct reference, A-form (Dir-ref A); Direct reference, X-form (Dir-ref X); Rule given, A-form (Rul-giv A); Rule given, X-form (Rul-giv X).

A total of 60 college student volunteers from *E*'s two sections of Educational Psychology, taught in the Spring of 1957, formed the original sample. The *Ss* were divided into six equal groups through the use of a table of random numbers.

The groups were compared in age, sex, grade level, and scholastic aptitude, and the differences were judged to be insignificant. However, nine of the original group were eliminated either because the recordings were defective or because they failed to complete the experiment. This left a total of eight in each group except one which had 10 remaining. Two more *Ss* were eliminated from the group with 10 *Ss* by the use of a table of random numbers so as to make the groups equal in number. The remaining 48 *Ss* provided the data which were used in the analysis.

The procedure was to have each of the *Ss* attempt to learn the two rules to the point where he could verbalize the rules and apply them in the solution of three different problems in succession. Tested individually, the *Ss* were asked to "think aloud" during the learning period, and voice recordings were made.

Immediately following the learning period, a test consisting of 20 problems was given. Then, approximately four weeks later, the *Ss* were asked to solve two simple problems and to fill out a questionnaire on their process of thinking.

The learning problems and the test problems were reproduced on separate pieces of paper or cardboard so that they could be presented one at a time. The experimental procedure may best be described step-wise, as follows:

Step 1. General instructions. The following instructions were given to each *S*.

I am going to show you a series of addition problems one at a time, and you are to try to discover how to find the sum of each problem without adding in the usual manner.

Or, in the event the rule was to be given, the initial statement was as follows.

I am going to tell you how to find the sum of two kinds of number series without adding, and then I will give you some practice in using the rules.

The remaining general instructions were the same for each experimental group.

There are two types of problems, and there is a different rule for each of the two types.

At this point two sample problems were presented, as illustrated in Fig. 5.

The first type, called the Odd-numbers problem, consists of a series of odd numbers

beginning with one. Here is an example. (Point to the sample problem in the A-form.) Below the series you are given the number of figures in the series. In this case, there are five figures, so the "number" is five.

Now look at the problem on the right. This problem is the same as the one on the left except that the numbers are represented in rows of X's. Each row represents a number. Do you see? (Explained, if necessary.)

The second type problem is called the Constant-difference problem. This type consists of numbers which increase by a certain amount. The series may begin with any number and be of any length, but the numbers will always increase by a certain amount. In the example, the numbers increase by one. (Point to the sample problem in the A-form.) Again, on the right is the same problem in the X-form.

From this point the instructions were somewhat different for each treatment combination. For the two *no-help* groups the instructions were as follows.

See if you can discover the rule for the Odd-numbers type problem first. Keep in mind that you are trying to discover a way of finding the sum of this type problem without adding in the usual manner. The problems I show you will all be examples of this type problem in the X-form (or conventional form).

At this point the first "learning" problem was presented. The learning problems appeared the same as the sample problem in Fig. 5, but were all in the appropriate number form.

I want to record what you are thinking, so "think aloud" as much as possible. Tell me what you are thinking, in other words.

After *S* learned the rule to the Odd-numbers type problem, the Constant-difference problem was introduced by simply saying, "Now let's see if you can discover the rule to the Constant-difference type problem."

The specific instructions to the two *direct-reference* groups were as follows:

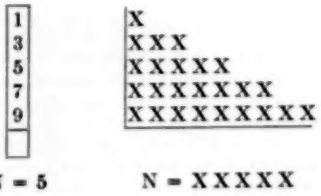

FIG. 5.

See if you can discover the rule for the Odd-numbers type problem first. Keep in mind that you are trying to discover a way of finding the sum of this type problem without adding in the usual manner. The problems I show you will all be examples of this type problem in the X-form (or conventional form).

In order to help you discover the rules, these examples provide you with additional hints. I'll show you these one at a time and explain the hints to you.

The first problem was presented at this point. Examples of the learning problems used with the Dir-ref X and Dir-ref A groups are illustrated in Fig. 1 and 2, respectively.

At this point, the instructions to the Dir-ref A group continued as follows.

The problem will always be inside the box, and the hint will be to the right of the box where you see the column of fours.

Notice that there are four 4's in the column, the same number of figures as there are in the problem series. Also notice that when you add the two columns, you get the same sum. Does that give you any ideas?

I want a record of what you are thinking, so "think aloud" as much as possible. Tell me what you are thinking, in other words.

After *S* learned the rule to the Odd-numbers type problem, the Constant-difference problem was introduced in the following manner:

Now let's see if you can discover the rule for the Constant-difference type problem. Again, the problem will always be inside the box, and the hint will be to the right of the box.

Essentially what we have done here is to take the problem series, invert it, and add across in the manner indicated. When we do this, you notice that we always get the same number, in this case 7. In other words, when you add the first number to the last number, you get 7; when you add the second number to the next-to-the-last, you get 7, and so on.

Also notice that the sum of the first column to the right is the same as that of the problem series, and the sum of the second column, the column of 7's, is twice the sum of the problem series. Now, does that suggest to you any ideas?

Starting again at the point at which the first problem is presented, the instructions to the Dir-ref X group continued as follows.

In these examples, the lines enclosing some of the X's and the dashes in a box are the hints. Notice that the number of dashes in the box is the same as the number of X's remaining outside. In other words, the total number of X's in the problem can be rearranged in the form of the box. Does this give you any ideas? Tell me what you are thinking. "Think aloud," in other words.

After the Odd-numbers rule was discovered, the Constant-difference problem was introduced as follows.

Now let's see if you can discover the rule for the Constant-difference type problem. Again the lines and dashes are your hints. Notice in this case that the number of dashes equals the number of X's. Essentially what we have done here is to invert the problem series and attach it to the original in the manner indicated to form the box. In other words, when you add the first row to the last you get 7; when you add the second row to the next-to-the-last, you get 7, and so on. In the end you have twice as many spaces within the box as there are X's in the original problem. Does that give you any ideas?

For the *rule-given* groups, the instructions were as follows:

The problems I will show you first are all examples of the Odd-numbers type problem in the X-form (or conventional form). The rule is to square the number of figures in the series to get the sum. The number of figures is 4. Four squared, or 4 times 4, is 16, which is the sum.

Now I will show you some more examples of this type problem. See if you can apply the rule correctly to each. Tell me what you are thinking.

After learning the Odd-numbers rule to the established criterion of three successful applications in succession, the Constant-difference type problem was introduced as follows.

Now let's see if you can learn the rule to the Constant-difference type problem. The rule is to add the first and the last numbers in the series, multiply by the number of figures in the series, and then divide by two. With the first example, add the first and the last numbers (2 and 5 are 7), multiply by the number (4 times 7 is 28), and divide by two (28 divided by 2 is 14). Fourteen is the correct sum. Now you practice the rule on these other examples.

Step 2. The learning period. The learning period was considered to begin immediately after the instructions in Step 1 were given, and to end immediately after the *S* successfully applied an acceptable rule to three problems in succession. An acceptable rule was defined as one which could be used in the solution of the problems as presented in the learning period, even though it might be inadequate for use with the problems as presented in the test period. For example, a rule which was frequently considered is to multiply the middle number in the series by the number of figures. This was adequate during the learning period because the problems included all the numbers; however, the test problems included only the first three numbers and the last of the series, which made the rule awkward to apply. This criterion forced *E* to deviate slightly from the experimental procedure in some cases, as will be described below.

Six different problems were used in the learning period, and they were basically

the same for each group. The general procedure was to wait until *S* had developed a possible solution with one problem, then to give him the next problem, telling him to try it on another example. This procedure was continued until *S* was successful. *S* was permitted to manipulate the cards himself if he wished, to back-track, or to look at several problems simultaneously. Also, *E* repeated the instructions when asked, and answered questions regarding the task. Otherwise, *E* limited his remarks to those which were intended to give support and encouragement, such as, "Try your idea on the next problem," "You are doing very well," and "See if it works." In order to keep *S* thinking aloud, *E* would ask, "What are you thinking now?" whenever *S* was silent for more than 10 or 15 seconds.

In the event that *S* discovered a rule which appeared to work satisfactorily but which was not the intended rule, *E* first attempted to encourage him to continue searching by saying, "That's fine, now see if you can discover another rule." In the event that *S* perseverated, *E* would shift to the problems used during the first test period, and, as soon as *S* became aware of the inadequacy of his rule, would shift back again to the learning problems.

There were some individuals, particularly in the no-help groups, who were simply unable to discover the intended rules within the time period of 60 to 90 minutes scheduled for each *S*. In this event, the learning period was terminated.

Step 3. The first test period. Immediately following the learning period, *S* was given 20 problems to solve, five each of the following: Odd-numbers problems (A-form); Odd-numbers problems (X-form); Constant-difference problems (A-form); Constant-difference problems (X-form).

The 10 problems of each type were actually five problems presented once in the A-form and once in the X-form. The problems were grouped by type and numerical form and were clearly identified as such. The procedure was to present the Odd-numbers problems first, then the Constant-difference problems. Furthermore, the *S*s who learned the rules using problems in the A-form were presented the test problems in the A-form first.

Step 4. The retest period. A retest was given to all *S*s four to six weeks after the first test. The second test consisted of two problems which could most easily be solved by using the two rules which the *S*s attempted to learn during the learning period, and a series of questions on their process of thinking. The problems and questions were completed under supervision, all in one day. The instructions were to record all scratchwork in spaces provided for this purpose, and to write the answers to the questions as clearly and completely as possible. The problems were as follows: (*a*) What is the sum of the first 35 odd numbers, and (*b*) what is the sum of *all* the first 35 numbers? The questions which followed each problem were the following:

> Did you add the first 35 (odd) numbers to get your answer? (Yes or no) If your answer is no, explain how you obtained your answer.
>
> Did you try to recall the rule you learned (or attempted to learn) under our direction several weeks ago? (Yes or no)
>
> Were you successful in recalling the rule? (Yes or no)
>
> Describe how you recalled or attempted to recall the rule.

Analysis of the Results

Not all *S*s succeeded in learning the two rules as stated above to the established criterion. Some of those who failed did manage to discover other workable variations of the rules which were judged to be acceptable. Others, however, failed to learn any workable rule for one of the two tasks before the practical limitations of time forced *E* to terminate the learning period.

TABLE 1

NUMBER WHO LEARNED EACH RULE (OR AN ACCEPTABLE VARIATION) TO CRITERION DURING THE LEARNING PERIOD

Rule[a]	No-help		Dir-ref		Rul-giv		All treatments
	A	X	A	X	A	X	
Odd-numbers rule							
1. n^2	7	6	8	7	8	8	44
2. m^2	1	0	0	0	0	0	1
3. mn	0	1	0	0	0	0	1
4. $[\frac{1}{2}(a + l)]^2$	0	0	0	1	0	0	1
5. none	0	1	0	0	0	0	1
All rules	8	8	8	8	8	8	48
Constant-difference rule							
1. $\frac{1}{2}n(a + l)$	1	0	7	7	8	8	31
2. mn	1	1	0	0	0	0	2
3. none	6	7	1	1	0	0	15
All rules	8	8	8	8	8	8	48

[a] In this column, n = number of figures in series; m = mean, or median of series; a = first number, and l = last number in series.

Table 1 shows the number of those who learned the rules and each of the acceptable variations, and those who failed to learn any workable rule in each case.

Those who failed to learn an acceptable rule were retained in the experiment and retested along with the others four weeks later, in anticipation of the possibility that their performance on the retest problems and their responses to the retest questionnaire would reveal useful information for further research. This indeed proved to be the case.

The purpose of the 20-item test given immediately after the learning period was to detect differences among the *Ss* in their achievement, if any. All *Ss* who learned the stated rules to criterion scored perfectly on the test, and those who failed to learn the acceptable rules were, of course, unable to solve the problems without adding.

On the retest given four weeks later, the datum of primary interest was the method used in solving the problems, not whether or not the correct answer was provided. Table 2 shows the methods used on the retest and the number of *Ss* in each group who used each method.

A comparison of Table 2 with Table 1 reveals the changes in the learned methods and the methods used on the retest, but does not indicate whether or not the methods that were used on the retest were the same as those which were developed during the learning period. It is apparent that there were at least three possibilities: (*a*) the same method was used, (*b*) a different method was used, or (*c*) the problem was solved by simple addition, which would indicate complete failure to recall or transfer the rule if learned. Table 3 shows the number of *Ss* in each of the teaching treatment groups who, on the retest, used the same method, some other method than that which they had learned, or simple addition. The *Ss* who failed to learn any acceptable rule in the learning period were placed in the "added" category if they added on the retest, and in the "other" category if they attempted some other procedure on the retest.

The results presented above fail to support the hypothesis that the *Ss* whose attention is directed to the relationships which are essential to understanding remember their learning longer and transfer it more effectively than do the *Ss* of the other two treatment groups (Hypothesis 3, above). The prediction was that the direct-reference group would be superior. Instead, although the obtained differences are not highly reliable, they are consistent with previously published data to the extent that they suggest that the independent discovery procedure is superior to learning by rote.

The reader's attention is directed to the fact that although 13 *Ss* in the no-help groups failed to learn an acceptable rule for the Constant-difference problem during the learning period, there were only four who added on the retest, and 10 who

TABLE 2

METHODS USED ON THE RETEST

Method[a]	No-help		Dir-ref		Rul-giv		All number treatments		All teaching treatments		
	A	X	A	X	A	X	A	X	No-help	Dir-ref	Rul-giv
Problem 1. Sum of the first 35 odd numbers											
1. n^2	6	6	2	2	5	1	13	9	12	4	6
2. $\frac{1}{2}n(a + l)$	0	1	0	1	0	0	0	2	1	1	0
3. $[\frac{1}{2}(a + l)]^2$	0	0	1	1	1	0	2	1	0	2	1
4. mn	2	0	0	0	0	0	2	0	2	0	0
							[b]		[b]		
5. Not acceptable	0	0	3	1	1	1	4	2	0	4	2
6. By addition	0	1	2	3	1	6	3	10	1	5	7
Chi square	Not applicable						1.39*		11.18**		
Problem 2. Sum of all the first 35 numbers											
1. $\frac{1}{2}n(a + l)$	3	2	3	4	4	0	10	6	5	7	4
2. mn	3	2	0	0	0	0	3	2	5	0	0
							[b]		[b]		
3. Not acceptable	1	1	3	2	1	0	5	3	2	5	1
4. By addition	1	3	2	2	3	8	6	13	4	4	11
Chi square	Not applicable						1.36*		3.86*		

[a] In this column, n = number of figures in series; m = mean, or median of series; a = first number, and l = last number in series.

[b] Frequencies above and below line combined for chi-square analysis.

* Not significant at the 0.05 level.

** Significant at the 0.05 level.

TABLE 3

CORRESPONDENCE OF METHODS USED ON THE RETEST WITH THOSE LEARNED DURING THE LEARNING PERIOD

Method	No-help		Dir-ref		Rul-giv		All number treatments		All teaching treatments		
	A	X	A	X	A	X	A	X	No-help	Dir-ref	Rul-giv
Problem 1. Sum of the first 35 odd numbers											
Same	6	5	2	3	5	1	13	9	11	5	6
Added	0	1	2	3	1	6	3	10	1	5	7
Others	2	2	4	2	2	1	8	5	4	6	3
Chi square	Not applicable						3.48*		8.21*		
Problem 2. Sum of all the first 35 numbers											
Same	3	2	3	3	4	0	10	5	5	6	4
Added	1	3	2	2	3	8	6	13	4	4	11
Other	4	3	3	3	1	0	8	6	7	6	1
Chi square	Not applicable						3.03*		9.99**		

* Not significant at the 0.05 level.

** Significant at the 0.05 level.

TABLE 4

NUMBER WHO WERE JUDGED TO BE COGNIZANT OF EACH TYPE OF RELATIONSHIP DURING THE LEARNING PERIOD

Type Relationship	No-help		Dir-ref		Rul-giv		All Treatments
	A	X	A	X	A	X	
	Odd-numbers rule						
Arithmetical	1	1	1	0	0	0	3
Geometrical	0	2	0	4	0	1	7
Both	0	0	0	1	0	0	1
Neither	7	5	7	3	8	7	37
	Constant-difference rule						
Arithmetical	3	1	6	0	0	0	10
Geometrical	0	0	0	2	0	0	2
Both	0	0	0	0	0	0	0
Neither	5	7	2	6	8	8	36

used acceptable methods. In the case of the other two teaching treatment groups, the number who added increased, and the number who used acceptable rules decreased rather markedly. This finding suggests that as a result of their experience during the learning period, the *Ss* in the no-help groups were motivated to continue learning afterwards and those treated otherwise were not.

The superiority of the discovery procedure may be better explained in terms of motivation, then, than in terms of understanding. This suggestion is given added support by the following data.

In an attempt to determine the number of *Ss* in each experimental group who actually became cognizant of the arithmetical, the geometrical, or both types of relationship, the typed transcriptions of the voice recordings made during the learning period were carefully examined. Any *S* was judged to be cognizant of a relationship if he verbalized it at any point in the process of his learning. As a result of *E*'s continued efforts to stimulate each *S* to think aloud during the learning period, very complete records were obtained. In addition, each *S* who failed to volunteer an explanation of the rule was asked to do so at the end of the period. In most cases of the latter type, the *S*'s response was that he could not provide an explanation. Table 4 shows the results of this analysis.

Most striking is the fact that only one-quarter of the entire group of 48 *Ss* was judged to be cognizant of one or both types of relationship to each of the two rules. The small number of *Ss* precludes any statistical check of the reliability of the obtained difference in the experimental groups. However, accepting the obtained data as reliable, the results appear to support the prediction that the direct-reference treatment would produce the greatest incidence of understanding. The next highest frequency of understanding was in the no-help group, and even in the rule-given group there was at least one *S* who understood the odd-numbers rule. Furthermore, with only two exceptions, both in the no-help group, the type of relationship discovered corresponded to the type that was predicted from the number treatment. When the problems were presented in the A-form, the arithmetical relationship was discovered, and when the X-form was used, the geometrical relationship was discovered. The data, therefore, tend to support Hypotheses 2 and 4.

The data presented in Table 4 do not preclude the possibility, however, that the *Ss* who were not cognizant of the relationships at the end of the learning period subsequently did become cognizant of them. The most adequate test of the efficacy of understanding in learning is to determine how those who actually verbalized the relationships performed on the retest problems. Table 5 shows the number of those who, on the retest, used the same rule they learned, those who simply added, and those who used some other method than that learned.

Again, the dearth of data does not per-

TABLE 5

METHODS USED ON RETEST OF THOSE WHO WERE JUDGED TO BE COGNIZANT OF THE ESSENTIAL RELATIONSHIPS

Methods	No-help		Dir-ref		Rul-giv		All Groups
	A	X	A	X	A	X	
	Problem 1						
Same Rule	1	3	0	2	0	0	6
Added	0	0	0	1	0	1	2
Others	0	0	1	2	0	0	3
All Methods	1	3	1	5	0	1	11
	Problem 2						
Same Rule	2	1	3	0	0	0	6
Added	1	0	0	2	0	0	3
Others	0	0	3	0	0	0	3
All Methods	3	1	6	2	0	0	12

mit a conclusive test of the first hypothesis. However, since only about one-half of those who learned acceptable rules with understanding during the learning period succeeded in retaining and transferring their learning four weeks later, the data do not offer very conclusive support of the meaning theory.

DISCUSSION

Perhaps the data which have been presented are sufficient to suggest the inadequacy of the "meaning" explanation, but they fail to communicate the impression of markedly increased motivation or interest in the task which characterized many of the *Ss* in the no-help group as compared with many of those in the other groups. The difference in motivation was reflected in their written comments on the retest and verbal reports to *E*. For example, one *S* in the no-help group reported that he was so intrigued with his success in discovering the rules that he told his friends of his experience and tested their ability. Others in the no-help group who failed to discover the rule told of their efforts to learn the rule, even going so far as to look up the algebraic formula in the library. On the other hand, one *S* in the rule-given group complained that *E* had not instructed him to remember the rules, so he had promptly forgotten them. Others in the rule-given and direct-reference groups commented on their confusion during the learning period in explaining their inability to recall the rules. The evaluation of the *Ss*' comments was admittedly highly subjective, but it may better suggest the nature of the motivation which is offered as the more adequate explanation for the superiority of the discovery procedure.

Of the various descriptive concepts of human motivation known to this present writer, Allport's concept of functional autonomy best describes the motivation that was developed by those in the no-help group. Whereas, at the beginning of the learning period, all the participants were assumed to be alike in that they were motivated primarily by such extrinsic factors as the approval of *E*, at the end of the learning period, the motivation of many of those in the no-help group appeared to be independent of the experimental or instructional situation. Presumably, the motivating power is of the type that lies in acquired interest or ego involvement in a task, and develops to the extent that the individual relies on his own cognitive capacities in learning.

The findings place the teacher in somewhat of a dilemma. He must decide which is the more important outcome of learning, (*a*) maximum understanding, or (*b*) maximum motivation to continue learning. However, further research on the problem may reveal ways of directing the discovery process without inhibiting the development of autonomous motivation. Teachers should continue to strive to guide the learning of their students but should refrain from giving answers directly.

Finally, the findings suggest that more consideration be given to the influence of the stimulus materials in directing the

thought processes of the learner. Particularly recommended is a careful study of the educational "toys" that are used in some primary classrooms to develop number concepts, and the practice exercises in workbooks and on worksheets which are intended to direct the learner's attention to relationships.

Summary and Conclusions

An experiment was performed with college students in order to test the premise that learning by independent discovery is superior to learning with direction because the learning is more meaningful in the former case. "Meaningful learning" was used in the cognitive sense of understanding or organization. The *Ss* were required to learn two arithmetical tasks to a common criterion. A two-factorial design was used which involved three "teaching" treatments (no-help, direct-reference, and rule-given), and two "number" treatments (A-form and X-form). Approximately four weeks later a retest was administered which was designed to test the ability to recall and apply the learned rules to a somewhat different type of problem from that used in the learning period. The *Ss*' self-reports of their learning process and methods used on the retest were analyzed.

The conclusion is that the superiority of the discovery procedure of learning over procedures of learning with external direction is not adequately explained in terms of "meaningful learning." The results of this experiment suggest that when the learner is forced to rely on his own cognitive capacities, it is more likely that he will become motivated to continue the learning process or to continue practicing the task after the learning period. Consequently, the learning becomes more permanent and is more effectively transferred than when the learner is not so motivated.

REFERENCES

1. Brownell, W. A., & Moser, H. E. Meaningful versus mechanical learning: A study in Grade III subtraction. *Duke Univer. Res. Stud. in Educ.*, 1949, No. 8.
2. Craig, R. C. *The transfer value of guided learning.* New York: Teachers Coll., Columbia Univer., 1953.
3. Craig, R. C. Directed versus independent discovery of established relations. *J. educ. Psychol.*, 1956, **47**, 223–234.
4. Katona, G. *Organizing and memorizing: Studies in the psychology of learning and teaching.* New York: Columbia Univer. Press, 1940.
5. McConnell, T. R. Discovery vs. authoritative identification in the learning of children. *Univer. Iowa Stud. in Educ.*, 1934, **9**, No. 5, 11–62.
6. Swenson, Esther, Anderson, G. L., & Stacey, C. L. *Learning theory in school situations.* Minneapolis: Univer. Minnesota Press, 1949.
7. Thiele, C. L. The contribution of generalization to the learning of the addition facts. *Teach. Coll. Contrib. Educ.*, 1939, No. 763.

Received July 11, 1958.

Again Available

SCHOOL PSYCHOLOGISTS AT MID-CENTURY

Edited by

NORMA E. CUTTS

The Report of the Thayer Conference, published in 1955, covered an intensive study of the development of school psychology. Because of continuing demand, the Report has been reprinted.

230 pages. Price $2.75

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Cole-Bruce

EDUCATIONAL PSYCHOLOGY

REVISED EDITION

This well-known textbook has been thoroughly revised with the help of critiques prepared by users of the first edition. The basic organization is retained, but many chapters have been entirely rewritten to make room for new material—e.g., a completely new treatment of measurement in education.

The authors provide a scholarly account of the psychology of the individual, give careful attention to child development, constantly relate psychological principles to classroom teaching, and discuss fully environmental influences on both child and teacher.

All references, readings, and film lists have been carefully checked and brought up to date. The book has a new format, with larger page size and more illustrations. Workbook and tests in preparation.

Write for College Catalog

WORLD BOOK COMPANY

Yonkers-on-Hudson, New York

Chicago, Boston, Atlanta, Dallas, Berkeley